



UNIVERSITÀ DEGLI STUDI DI MACERATA

**CORSO DI DOTTORATO DI RICERCA IN
Umanesimo e Tecnologie**

CICLO XXXVIII

TITOLO DELLA TESI

**Studio sulla traduzione automatica del linguaggio scientifico tedesco in italiano
e sulla valutazione della sua qualità**

SUPERVISORE DI TESI
Chiar.ma Prof.ssa Antonella Nardi

DOTTORANDA
Dott.ssa Federica Lattanzio

CO-SUPERVISORE DI TESI
Chiar.ma Prof.ssa Raffaella Merlini

COORDINATORE
Chiar.mo Prof. Roberto Lambertini

ANNO 2026



Indice

Introduzione	1
1 Le tecnologie per la traduzione assistita e automatica	6
1.1 Concetti e termini.....	6
1.2 Traduzione assistita	9
1.2.1 CAT tools	10
1.2.2 Vantaggi e svantaggi nell'uso dei CAT tools	14
1.3 Traduzione automatica	15
1.3.1 Principali architetture dei sistemi di TA.....	17
1.3.1.1 Dagli approcci linguistici agli approcci statistici.....	17
1.3.1.2 Traduzione automatica neurale	21
1.3.1.2.1 <i>Word embedding</i>	23
1.3.1.2.2 Architetture delle reti	24
1.3.1.2.3 <i>Training</i>	28
1.3.2 Traduzione automatica adattiva	29
1.4. <i>Augmented Translation</i>	31
1.5 Intelligenza artificiale e <i>Large Language Model</i>	33
2 Valutazione della qualità della TA	36
2.1 Concetti e definizioni	36
2.2 Metriche di valutazione della qualità	37
2.2.1 Tecniche di valutazione umana della TA	38
2.2.1.1 Metodi di valutazione <i>DEJ-based</i>	38
2.2.1.2 Metodi di valutazione <i>non-DEJ-based</i>	41
2.2.1.3 Vantaggi e svantaggi delle tecniche di valutazione umana della TA	43
2.2.2 Tecniche di valutazione automatica della TA	46
2.2.2.1 Metriche basate su traduzioni di riferimento	46
2.2.2.2 <i>Confidence</i> o <i>quality estimation</i>	48
2.2.2.3 Valutazione basata sui <i>checkpoints</i>	49
2.2.2.4 Vantaggi e svantaggi delle tecniche di valutazione automatica della TA	50
3 Post-editing	52
3.1 Concetti e definizioni	52
3.2 I livelli di <i>post-editing</i>	54

3.2.1	Linee guida per il <i>light</i> e il <i>full post-editing</i>	56
3.2.2	Necessità di aggiornamento delle linee guida per <i>light</i> e <i>full post-editing</i>	57
3.3.	<i>Post-editing</i> monolingue vs <i>post-editing</i> bilingue	59
3.4	ISO 18587: <i>post-editing</i> standard	61
3.5	Le competenze del <i>post-editor</i>	62
3.6	La misurazione dell' <i>effort</i> e la retribuzione del <i>post-editor</i>	65
3.7	Linee guida per la classificazione degli errori	67
3.8	Sfide e aspetti problematici del <i>post-editing</i>	69
4	La scrittura scientifica	72
4.1	I livelli di comunicazione nella scienza	72
4.2	Le forme del testo scientifico scritto	74
4.2.1	Tipologie di articolo scientifico	76
4.2.2	Componenti dell'articolo scientifico	77
4.3	Il linguaggio scientifico	78
4.3.1	Linguaggio specialistico e linguaggio scientifico	80
4.3.2	Anglicizzazione vs plurilinguismo	81
4.3.3	Le componenti del linguaggio scientifico	83
4.3.3.1	Il linguaggio quotidiano e il linguaggio specialistico	83
4.3.3.2	Il linguaggio scientifico quotidiano	86
4.3.4	Il tedesco come linguaggio scientifico	88
4.3.5	L'italiano come linguaggio scientifico	92
5	Studio empirico: traduzione con DeepL su Trados Studio	95
5.1	Materiali	95
5.2	Strumenti	97
5.2.1	DeepL Translator	97
5.2.1.1	Caratteristiche e componenti	97
5.2.1.2	Come funziona DeepL?	99
5.2.2	Trados Studio	101
5.2.2.1	Componenti essenziali di Trados Studio	102
5.2.2.2	Come funziona Trados Studio?	103
5.2.3	Sketch Engine	104
5.3	Fasi dello studio	108
5.3.1	Creazione dei primi progetti su Trados Studio	109

5.3.2 Interventi di <i>pre-editing</i>	112
5.3.3 Creazione dei secondi progetti su Trados Studio	113
5.3.4 Individuazione e annotazione degli errori.....	114
5.3.5 Tassonomia per Errori nella Traduzione Automatica Scientifica Tedesco-Italiana: TETASTI.....	115
5.3.5.1 Modelli di riferimento.....	115
5.3.5.2 Descrizione di TETASTI.....	120
5.3.5.2.1 Errori di Accuracy.....	122
5.3.5.2.2 Errori di <i>Fluency</i>	131
5.3.5.2.3 Errori di <i>Grammar</i>	138
5.6 <i>Post-editing</i>	141
6 Studio empirico: analisi dei risultati e discussione	145
6.1 Procedura adottata per l'analisi degli errori.....	145
6.2 Progetto n. 1: <i>Informationen Deutsch als Fremdsprache</i>	148
6.2.1 Analisi degli errori	148
6.2.2 Analisi dell' <i>effort</i> temporale e tecnico	162
6.3 Progetto n. 2: <i>trans-kom</i>	165
6.3.1 Analisi degli errori	165
6.3.2 Analisi dell' <i>effort</i> temporale e tecnico	179
6.4 Progetto n. 3: <i>Interculture journal</i>	182
6.4.1 Analisi degli errori	182
6.4.2 Analisi dell' <i>effort</i> temporale e tecnico	196
6.4 Riflessione conclusiva.....	199
7 Studio sperimentale: traduzione con un sistema di TA addestrato e <i>post-editing</i>	204
7.1 Strumenti.....	204
7.1.1 Transkribus.....	205
7.1.2 LF Aligner.....	206
7.1.3 OPUS-CAT MT Engine	209
7.2 Prima prova di <i>fine-tuning</i>	212
7.2.1 Costituzione del corpus.....	213
7.2.2 <i>Fine-tuning</i> manuale	214
7.3 Seconda prova di <i>fine-tuning</i>	217
7.3.1 Costituzione del corpus.....	218

7.3.1.1 Scelta dei testi	218
7.3.1.2 <i>Data cleaning</i> manuale	219
7.3.1.3 <i>Data cleaning</i> automatico	221
7.3.1.4 Traduzione, segmentazione e allineamento	223
7.3.2 <i>Fine-tuning</i> assistito	224
7.3.3 Valutazione automatica	228
7.3.4 Traduzione, annotazione degli errori e <i>post-editing</i>	231
8 Testi tradotti con Deepl e il sistema di TA addestrato: un confronto	233
8.1 Testo per il progetto n. 1 (8.454 <i>tokens</i>).....	234
8.1.1 Analisi degli errori	234
8.1.2 Analisi dell' <i>effort</i> temporale e tecnico	250
8.2 Testo per il progetto n. 2 (8.484 <i>tokens</i>).....	253
8.2.1 Analisi degli errori	253
8.2.2 Analisi dell' <i>effort</i> temporale e tecnico	271
8.3 Testo per il progetto n.3 (6.120 <i>tokens</i>).....	273
8.3.1 Analisi degli errori	274
8.3.2 Analisi dell' <i>effort</i> temporale e tecnico	293
8.4 Riflessione conclusiva.....	296
9 Conclusioni e prospettive.....	301
9.1 Risultati della ricerca	301
9.2 Risposta alle domande di ricerca	306
9.3 Prospettive per ulteriori ricerche.....	309
Bibliografia	312
Sitografia.....	334
Allegati	337

Indice delle figure

Figura 1: Traduzione umana e traduzione automatica	7
Figura 2: Editor di testo di Trados Studio	13
Figura 3: Illustrazione di una rete neurale	22
Figura 4: Rete neurale ricorrente	25
Figura 5: Augmented Translation puts Linguists in the center	32
Figura 6: Ecosistema di linee guida del <i>post-editing</i>	58
Figura 7: Abilità e competenze per il <i>post-editing</i>	63
Figura 8: Lessico comune e lessico specialistico	84
Figura 9: Grado di specializzazione linguistica	85
Figura 10: Interfaccia di DeepL n.1	100
Figura 11: Interfaccia di DeepL n.2	101
Figura 12: <i>Work Pane</i> di Trados Studio	103
Figura 13: <i>Dashboard</i> di Sketch Engine	105
Figura 14: Funzione Concorde per la parola “ricerca”	105
Figura 15: Funzione Word Sketch per la parola “analisi”	107
Figura 16: Funzione <i>Thesaurus</i> per la parola “risultato”	108
Figura 17: Strumenti di controllo della qualità 3.0	110
Figura 18: Esempio di resa problematica con Trados Studio	111
Figura 19: Esempio n.1 <i>pre-editing</i>	112
Figura 20: Esempio n.2 <i>pre-editing</i>	113
Figura 21: Esempio n.1 di errore di segmentazione	113
Figura 22: Esempio n.2 di errore di segmentazione	113
Figura 23: Esempio n.1 segmentazione corretta	114
Figura 24: Esempio di punteggiatura superflua nel file Word	143
Figura 25: Esempio di punteggiatura aggiunta nell’editor di testo	143
Figura 26: Funzione Unisci segmenti in Trados Studio	144
Figura 27: Report di Post-Edit Compare: analisi dei segmenti – testo n. 6 dal progetto n. 2	146
Figura 28: Report di Post-Edit Compare: <i>Translation Modifications</i> e <i>Post-Edit Modification Analysis</i> – progetto n. 3	147
Figura 29: Esempio fascia di PEMp 100% dal testo n. 8 del progetto n. 1	163
Figura 30: Esempio fascia di PEMp 85-94% dal testo n. 8 del progetto n. 1	164
Figura 31: Esempio categoria di PEMp <i>New</i> dal testo n.8 del progetto n. 1	165

Figura 32: Esempio fascia di PEMp 100% dal testo n. 1 progetto n. 2	181
Figura 33: Esempio fascia di PEMp 85-94% dal testo n. 1 del progetto n. 2	181
Figura 34: Esempio categoria di PEMp <i>New</i> dal testo n. 4 del progetto n. 2	182
Figura 35: Esempio fascia di PEMp 95-99% dal testo n. 2 del progetto n. 3	197
Figura 36: Esempio fascia di PEMp 85-94% dal testo n. 1 del progetto n. 3	198
Figura 37: Esempio della categoria di PEMp <i>New</i> dal testo n. 7 per progetto n. 3	198
Figura 38: Visualizzazione <i>side-by-side</i> di Transkribus	206
Figura 39: Finestra per la selezione del formato dei file.....	207
Figura 40: Finestra per l'impostazione delle lingue	207
Figura 41: Finestra per la selezione dei file	208
Figura 42: Finestra per opzioni di revisione	208
Figura 43: <i>Graphical editor</i> di LF Aligner	209
Figura 44: Schermata per la selezione di modelli in OPUS-CAT MT Engine.....	211
Figura 45: Schermata dei modelli installati in OPUS-CAT MT Engine.....	211
Figura 46: Schermata di traduzione in OPUS-CAT MT Engine.....	212
Figura 47: Problematica spazi aggiuntivi tra le parole	220
Figura 48: Problematica interruzione di riga nel file .txt.....	220
Figura 49: Problematica interruzione di riga nel file .pdf.....	221
Figura 50: Uso del punto interrogativo nell'elisione	221
Figura 51: Uso del punto interrogativo nel troncamento	221
Figura 52: Problematica suddivisione di parole.....	222
Figura 53: Codice in Python per correggere l'uso improprio di “-”	222
Figura 54: Problema di segmentazione in LF Aligner	223
Figura 55: Schermata modelli installati in OPUS-CAT MT Engine.....	225
Figura 56: Schermata fine-tuning in OPUS-CAT MT Engine	226
Figura 57: Schermata fine-tuning in OPUS-CAT MT Engine	227
Figura 58: Schermata Fine-tuning progress in OPUS-CAT MT Engine.....	228
Figura 59: Andamento di BLEU <i>score</i>	229
Figura 60: Andamento della <i>training</i> e <i>validation loss</i>	230

Indice delle tabelle

Tabella 1: Errori di <i>Accuracy</i>	122
Tabella 2: Errori di <i>Fluency</i>	132
Tabella 3: Errori di <i>Grammar</i>	138
Tabella 4: Analisi quantitativa degli errori per progetto n. 1	149
Tabella 5: Distribuzione delle fasce di PEMp per progetto n. 1	163
Tabella 6: Analisi quantitativa degli errori per progetto n. 2	166
Tabella 7: Distribuzione delle fasce di PEMp per progetto n. 2	180
Tabella 8: Analisi quantitativa degli errori per progetto n. 3	183
Tabella 9: Distribuzione delle fasce di PEMp per progetto n. 3	197
Tabella 10: Testi per il <i>training set</i>	219
Tabella 11: Analisi quantitativa comparativa degli errori per il testo del modello <i>fine-tuned</i> n. 1 ..	235
Tabella 12: Distribuzione comparativa delle fasce di PEMp per il testo del modello <i>fine-tuned</i> n. 1 ..	252
Tabella 13: Analisi quantitativa comparativa degli errori per il testo del modello <i>fine-tuned</i> n. 2 ..	254
Tabella 14: Distribuzione comparative delle fasce di PEMp per testo del modello <i>fine-tuned</i> n. 2 ..	272
Tabella 15: Analisi quantitativa comparativa degli errori per il testo del modello <i>fine-tuned</i> n. 3 ..	274
Tabella 16: Distribuzione comparativa delle fasce di PEMp per testo del modello <i>fine-tuned</i> n. 3 ..	295
Tabella 1 (All. 2): Analisi quantitativa degli errori per il testo n. 1 del progetto n. 1	340
Tabella 2 (All.2): Distribuzione delle fasce di PEMp per testo n. 1 del progetto n. 1	341
Tabella 3 (All. 2): Analisi quantitativa degli errori per il testo n. 2 del progetto n. 1	342
Tabella 4 (All. 2): Distribuzione delle fasce di PEMp per testo n. 2 del progetto n. 1	343
Tabella 5 (All. 2): Analisi quantitativa degli errori per il testo n. 3 del progetto n. 1	345
Tabella 6 (All. 2): Distribuzione delle fasce di PEMp per testo n. 3 del progetto n. 1	346
Tabella 7 (All. 2): Analisi quantitativa degli errori per il testo n. 4 del progetto n. 1	347
Tabella 8 (All. 2): Distribuzione delle fasce di PEMp per testo n. 4 del progetto n. 1	348
Tabella 9 (All. 2): Analisi quantitativa degli errori per il testo n. 5 del progetto n. 1	350
Tabella 10 (All. 2): Distribuzione delle fasce di PEMp per testo n. 5 del progetto n. 1	351
Tabella 11 (All. 2): Analisi quantitativa degli errori per il testo n. 6 del progetto n. 1	352
Tabella 12 (All. 2): Distribuzione delle fasce di PEMp per testo n. 6 del progetto n. 1	353
Tabella 13 (All. 2): Analisi quantitativa degli errori per il testo n. 7 del progetto n. 1	355
Tabella 14 (All. 2): Distribuzione delle fasce di PEMp per testo n.7 del progetto n. 1	356

Tabella 15 (All. 2): Analisi quantitativa degli errori per il testo n. 8 del progetto n. 1	357
Tabella 16 (All. 2): Distribuzione delle fasce di PEMp per testo n. 8 del progetto n. 1	358
Tabella 17 (All. 2): Analisi quantitativa degli errori per il testo n. 9 del progetto n. 1	360
Tabella 18 (All. 2): Distribuzione delle fasce di PEMp per testo n. 9 del progetto n. 1	361
Tabella 1 (All. 3): Analisi quantitativa degli errori per il testo n. 1 del progetto n. 2.....	365
Tabella 2 (All. 3): Distribuzione delle fasce di PEMp per testo n. 1 del progetto n. 2.....	366
Tabella 3 (All. 3): Analisi quantitativa degli errori per il testo n. 2 del progetto n. 2.....	367
Tabella 4 (All. 3): Distribuzione delle fasce di PEMp per testo n. 2 del progetto n. 2.....	368
Tabella 5 (All. 3): Analisi quantitativa degli errori per il testo n. 3 del progetto n. 2.....	370
Tabella 6 (All. 3): Distribuzione delle fasce di PEMp per testo n. 3 del progetto n. 2.....	371
Tabella 7 (All. 3): Analisi quantitativa degli errori per il testo n. 4 del progetto n. 2.....	372
Tabella 8 (All. 3): Distribuzione delle fasce di PEMp per testo n. 4 del progetto n. 2.....	373
Tabella 9 (All. 3): Analisi quantitativa degli errori per il testo n. 5 del progetto n. 2.....	375
Tabella 10 (All. 3): Distribuzione delle fasce di PEMp per testo n. 5 del progetto n. 2.....	375
Tabella 11 (All. 3): Analisi quantitativa degli errori per il testo n. 6 del progetto n. 2.....	377
Tabella 12 (All. 3): Distribuzione delle fasce di PEMp per testo n. 6 del progetto n. 2.....	378
Tabella 13 (All. 3): Analisi quantitativa degli errori per il testo n. 7 del progetto n. 2.....	379
Tabella 14 (All. 3): Distribuzione delle fasce di PEMp per testo n. 7 del progetto n. 2.....	380
Tabella 15 (All. 3): Analisi quantitativa degli errori per il testo n. 8 del progetto n. 2.....	382
Tabella 16 (All. 3): Distribuzione delle fasce di PEMp per testo n. 8 del progetto n. 2.....	383
Tabella 17 (All. 3): Analisi quantitativa degli errori per il testo n. 9 del progetto n. 2.....	385
Tabella 18 (All. 3): Distribuzione delle fasce di PEMp per testo n. 9 del progetto n. 2.....	386
Tabella 1 (All. 4): Analisi quantitativa degli errori per il testo n. 1 del progetto n. 3.....	389
Tabella 2 (All. 4): Distribuzione delle fasce di PEMp per testo n. 1 del progetto n. 3.....	390
Tabella 3 (All. 4): Analisi quantitativa degli errori per il testo n. 2 del progetto n. 3.....	391
Tabella 4 (All. 4): Distribuzione delle fasce di PEMp per testo n. 2 del progetto n. 3.....	392
Tabella 5 (All. 4): Analisi quantitativa degli errori per il testo n. 3 del progetto n. 3.....	394
Tabella 6 (All. 4): Distribuzione delle fasce di PEMp per testo n. 3 del progetto n. 3.....	395
Tabella 7 (All. 4): Analisi quantitativa degli errori per il testo n. 4 del progetto n. 3.....	396
Tabella 8 (All. 4): Distribuzione delle fasce di PEMp per testo n. 4 del progetto n. 3.....	397
Tabella 9 (All. 4): Analisi quantitativa degli errori per il testo n. 5 del progetto n. 3.....	399
Tabella 10 (All. 4): Distribuzione delle fasce di PEMp per testo n. 5 del progetto n. 3.....	400
Tabella 11 (All. 4): Analisi quantitativa degli errori per il testo n. 6 del progetto n. 3.....	401
Tabella 12 (All. 4): Distribuzione delle fasce di PEMp per testo n. 6 del progetto n. 3.....	402

Tabella 13 (All. 4): Analisi quantitativa degli errori per il testo n.7 del progetto n. 3	404
Tabella 14 (All. 4): Distribuzione delle fasce di PEMp per testo n. 7 del progetto n. 3	405
Tabella 15 (All. 4): Analisi quantitativa degli errori per il testo n. 8 del progetto n. 3	406
Tabella 16 (All. 4): Distribuzione delle fasce di PEMp per testo n. 8 del progetto n. 3	407
Tabella 17 (All. 4): Analisi quantitativa degli errori per il testo n. 9 del progetto n. 3	409
Tabella 18 (All. 4): Distribuzione delle fasce di PEMp per testo n. 9 del progetto n. 3	410

Introduzione

Nel panorama contemporaneo, l'innovazione tecnologica riveste un ruolo centrale in numerosi ambiti, esercitando un'influenza sempre più significativa anche nel campo degli studi traduttivi. Il suo impatto è particolarmente evidente nell'ampia diffusione della *Computer Assisted Translation* (CAT), ossia il procedimento di traduzione in cui il traduttore è assistito da programmi informatici, comunemente chiamati CAT tools, ormai indispensabili per l'esercizio della professione. Parallelamente al consolidamento della traduzione umana assistita da computer, si è osservato lo sviluppo della traduzione automatica assistita dall'uomo (*Human-Aided Machine Translation*, HAMT) (Hutchins e Somers 1992), che prevede l'intervento del traduttore a supporto del sistema di traduzione automatica in due modalità differenti: il *pre-editing* e il *post-editing*. Il *pre-editor* interviene sul testo di partenza prima della traduzione automatica, con l'obiettivo di facilitare l'elaborazione da parte del sistema. Diversamente, il *post-editor* corregge il testo generato dalla macchina con l'obiettivo di migliorarne la qualità (Scarpa 2008). I due interventi sopracitati rientrano in quella che più genericamente viene definita traduzione automatica (di seguito TA). In quest'ottica, la traduzione si configura come un'attività di interazione tra essere umano e macchina (Alonso e Calvo 2015; O'Brien 2012).

Il progresso tecnologico e la diffusione capillare della TA tra i *Language Service Providers* (LSP) hanno trasformato in modo sostanziale non solo la pratica professionale, ma anche la percezione stessa del ruolo del traduttore. Come evidenzia Sakamoto (2019), l'intervento umano tende oggi a concentrarsi prevalentemente nella fase di *post-editing*, con un conseguente ridimensionamento della componente decisionale e creativa che tradizionalmente caratterizzava la professione del traduttore. Tale riconfigurazione del ruolo traduttivo ha suscitato reazioni contrastanti, con atteggiamenti di maggiore apertura da parte di alcuni, ma anche paura, disagio e resistenza da parte di altri (Cadwell et al. 2018). Non sorprende quindi che, come riassume Sakamoto (2019: 201), molti traduttori “feel their professional skills and identities are sidelined by technology”. La studiosa suggerisce che la tecnologia non solo ristrutturava pratiche e ruoli, ma ridistribuisce anche il valore simbolico associato al lavoro traduttivo. In questo contesto, la tensione tra umanità e automazione non riguarda soltanto l'efficienza del processo traduttivo, ma coinvolge anche la dimensione identitaria e professionale del traduttore contemporaneo.

Questo scenario ha alimentato una diffusa *automation anxiety* (Vieira 2020), ossia la percezione di una professione progressivamente sotto pressione a causa dell'automazione, della riduzione dei compensi e della competizione globalizzata. Tale tendenza trova riscontro anche nei dati più recenti. Secondo l'ELIS (*European Language Industry Survey*) 2025, più del 50% delle

aziende di servizi linguistici e dei professionisti indipendenti dichiara di utilizzare la TA in oltre la metà del proprio lavoro. Tuttavia, il rapporto delinea un quadro di profonda e diffusa incertezza: il 53% dei traduttori freelance manifesta dubbi sulla sostenibilità futura della propria carriera e il 23% ipotizza l'abbandono dell'attività autonoma. Tale pessimismo appare strettamente correlato a una percezione negativa dell'intelligenza artificiale e della TA, considerate uno sviluppo positivo solo dal 13% dei professionisti indipendenti. Ciononostante, quest'ultimo gruppo esprime una maggiore fiducia in una progressiva maturazione della consapevolezza dei clienti, i quali, riconoscendo i limiti intrinseci di queste tecnologie, potrebbero tornare a privilegiare la qualità e la competenza dei servizi linguistici umani (ELIS 2025).

Tali dinamiche suggeriscono come la riduzione del dibattito sulla TA all'opposizione "essere umano o macchina" risulti metodologicamente poco produttiva. Appare piuttosto necessario adottare una prospettiva integrata che combini i vantaggi e gli svantaggi di entrambe le parti: la TA consente di ottenere traduzioni in tempi rapidi e con costi contenuti, ma non può eguagliare la qualità, la sensibilità linguistica e la capacità interpretativa del traduttore umano. In questa prospettiva, il *post-editing* non si configura solo come una fase operativa, ma come un terreno d'incontro, in cui competenza umana ed efficienza tecnologica possono integrarsi in modo complementare. La presente ricerca si propone, pertanto, di indagare se tale equilibrio sia effettivamente possibile e in che misura la collaborazione tra traduttore e macchina possa tradursi in una sinergia efficace. Il mercato del lavoro, in questo senso, sembra già fornire una risposta concreta: la domanda di professionisti specializzati nel *post-editing* è in costante crescita. Come evidenziato dal rapporto ELIS 2025, questa attività rappresenta ormai circa il 27% del fatturato complessivo del settore, confermandone il ruolo sempre più centrale nell'ecosistema traduttivo contemporaneo.

In quest'ottica, appare doveroso distinguere il profilo del traduttore da quello del *post-editor*, evitando di sovrapporre figure professionali che, pur condividendo una comune matrice traduttologica, afferiscono a domini operativi distinti. Il profilo professionale del *post-editor* richiede infatti non solo le competenze di traduttologia proprie del traduttore "tradizionale", bensì anche nozioni di linguistica computazionale, comprensione dei paradigmi di *machine learning* e una conoscenza approfondita delle architetture di TA. A tali competenze tecniche si affiancano abilità critiche e capacità di valutazione oggettiva, indispensabili per operare in modo consapevole. Il *post-editor* deve essere in grado di motivare i propri interventi e riconoscere quando questi siano effettivamente necessari, distinguendo le decisioni che prenderebbe come traduttore da quelle specifiche del *post-editor*. Tali indicazioni confermano che il passaggio dalla traduzione tradizionale al *post-editing* richieda una formazione adeguata e non possa essere affidato

all'intuizione. Lo studio di Scansani et al. (2019) conferma tale necessità, dimostrando inoltre che una formazione specifica sulle tecnologie basate sulla TA prima dell'ingresso nel mercato professionale consenta ai traduttori più giovani di essere meno influenzati dai pregiudizi tradizionalmente associati al *post-editing*.

La crescente complessità del lavoro di *post-editing* ha reso imprescindibile disporre di linee guida specifiche per l'annotazione degli errori nei testi tradotti automaticamente, intese come strumenti concepiti per coppie linguistiche, tipologie testuali o domini specialistici determinati. Negli ultimi anni sono state sviluppate diverse linee guida: alcune presentano un carattere generale, come MQM (*Multidimensional Quality Metrics*) framework (Lommel, Uszkoreit, et al. 2014; Lommel et al. 2013); mentre altre risultano specifiche, come SCATE (*Smart Computer-aided Translation Environment*) MT error taxonomy sviluppata da Tezcan et al. (2017) e applicata esclusivamente alla coppia linguistica inglese-olandese. Il carattere generico delle linee guida comporta da un lato, il vantaggio di adattarsi a qualunque genere testuale, coppia linguistica e ambito specialistico; dall'altro, può risultare limitante nella descrizione di fenomeni linguistici specifici di determinate coppie linguistiche o domini specialistici. Analogamente, anche le linee guida specifiche possono incontrare difficoltà nel classificare in modo adeguato errori che esulano dal contesto per cui sono state concepite.

Alla luce di tali considerazioni, emerge l'esigenza di ampliare e/o ridefinire le linee guida già esistenti o, laddove necessario, elaborare linee guida interamente nuove, ideate appositamente per il genere testuale, l'ambito specialistico e la coppia linguistica trattati. Questo è quanto si è verificato nello studio condotto da Tezcan et al. (2019), che ha previsto l'adattamento della tassonomia SCATE per l'annotazione di errori riscontrabili in un romanzo tradotto per la coppia linguistica inglese-olandese, introducendo nuove categorie di errori per descrivere più accuratamente le problematiche legate alle peculiarità dei testi presi in esame.

Nonostante i significativi progressi compiuti, permangono tuttavia aree ancora poco esplorate in questo ambito, tra cui l'applicazione di tali linee guida a generi testuali specifici, ad esempio il contributo scientifico. La presente proposta di ricerca si colloca in questa prospettiva, proponendosi di colmare tale lacuna.

Le motivazioni sottese all'indagine qui presentata sorgono dunque come risposta a svariate esigenze. In primo luogo, la già menzionata crescente rilevanza del *post-editing* sul mercato dei servizi linguistici rende necessario approfondirne l'applicazione nei diversi contesti specialistici. Come dimostrato dallo studio di Heiss e Soffritti (2018), i risultati del *post-editing* sono soggetti a variazioni significative in funzione dell'ambito di applicazione. Sebbene l'analisi dei due studiosi, condotta su un numero limitato di campioni appartenenti ai settori giuridico, accademico e tecnico,

non consenta generalizzazioni definitive, essa evidenzia l'esistenza di testi, come quelli altamente specialistici, che si prestano in misura minore rispetto ad altri a questo tipo di intervento. Da ciò deriva la scelta di indagare il genere testuale specifico del contributo scientifico nell'ambito specialistico della linguistica per la coppia tedesco-italiano, tenendo conto anche, per quanto emerge dalla disamina della letteratura, dell'assenza a oggi di studi dedicati a tali aspetti. Si avverte, inoltre, l'esigenza di redigere linee guida per l'annotazione di errori, per quanto noto non esistenti, che non siano di carattere generale, ma che siano calibrate sulle specificità linguistiche, terminologiche e strutturali proprie del linguaggio e del contributo scientifico. Pertanto, la decisione di dedicarsi al linguaggio scientifico (*Wissenschaftssprache*) non è una scelta casuale: esso rappresenta inoltre una dimensione linguistica estremamente complessa e questo è dovuto, come affermato da Ehlich (1996/2007), ai suoi tre elementi costituenti: il linguaggio comune, il linguaggio scientifico quotidiano (*alltägliche Wissenschaftssprache*) e la terminologia. Tale strutturazione rende particolarmente rilevante l'analisi della qualità dell'output dei sistemi di TA e del successivo intervento di *post-editing* nel presente studio.

Le domande di ricerca che si pongono con il presente elaborato sono le seguenti: quale livello di qualità presenta l'output grezzo generato dal sistema di TA? E conseguentemente: quanto lavoro è richiesto al *post-editor* per elaborare proposte di traduzione utili al raggiungimento di una qualità pubblicabile, ossia quella che Monti (2019) definisce come *outbound translation*?

Per rispondere a tali domande sono presi in esame contributi in lingua tedesca tradotti in italiano dal sistema di TA generico DeepL e successivamente sottoposti a una fase di individuazione e classificazione degli errori e, infine, di *post-editing*. In seguito all'analisi degli errori negli output della TA, sono state elaborate linee guida specifiche volte a ottimizzare il processo di *post-editing* e garantire una qualità del testo finale idonea alla pubblicazione. La redazione delle linee guida si è basata sugli errori individuati nelle traduzioni automatiche dei contributi scientifici presi in esame. I risultati di questo studio hanno mostrato che i sistemi di TA generici, come DeepL, non essendo addestrati specificatamente per la traduzione del linguaggio scientifico, non restituiscono costantemente risultati ottimali. Tale riscontro ha determinato l'integrazione nella ricerca di una fase sperimentale, volta a verificare se la qualità dell'output possa migliorare attraverso l'addestramento di un sistema di TA pre-addestrato (*fine-tuning*) per la traduzione di testi scientifici dal tedesco all'italiano su un corpus appositamente creato.

Il presente lavoro si compone di nove capitoli. Il primo delinea il quadro teorico di riferimento relativo alle tecnologie della traduzione, con particolare enfasi sulle principali evoluzioni della traduzione assistita e automatica. Il secondo capitolo approfondisce le metodologie di valutazione della qualità della TA, analizzando criticamente gli approcci di valutazione umana e

quelli di valutazione automatica. Il terzo capitolo è interamente dedicato alla fase di *post-editing*: si offre un'ampia panoramica sui concetti teorici e pratici fondamentali, presentando anche tassonomie e linee guida funzionali all'annotazione di errori. Il quarto capitolo tratta la scrittura scientifica, soffermandosi in primo luogo sul testo scientifico e successivamente sul linguaggio scientifico, evidenziandone gli aspetti rilevanti per la presente ricerca. Successivamente, il quinto capitolo è dedicato allo studio empirico che prevede la traduzione in italiano mediante il sistema di TA DeepL e il *post-editing* di un corpus composto da articoli scientifici tedeschi. Nello specifico si illustrano la metodologia, gli strumenti e i materiali impiegati, unitamente alle linee guida per la classificazione degli errori elaborate appositamente per questa indagine. Il sesto capitolo propone di analizzare i risultati emersi dall'annotazione e dal *post-editing* del suddetto corpus. Nel capitolo sette viene presentato lo studio sperimentale riguardante il *fine-tuning* di un sistema di TA pre-addestrato per la traduzione del linguaggio scientifico tedesco in italiano. Più nello specifico, in questo capitolo vengono illustrate le due prove di *fine-tuning* con i rispettivi strumenti impiegati e la metodologia adoperata. Successivamente, nel capitolo otto vengono esposti i risultati di un'analisi comparativa tra i testi tradotti con il modello *fine-tuned* e quelli prodotti dal sistema generico DeepL; l'intento è individuare le differenze qualitative e quantitative tra i due output e quindi tra i due sistemi. In chiusura, le conclusioni riportano una riflessione sugli studi condotti, dei risultati conseguiti e ulteriori spunti per future linee di ricerca.

Le tecnologie per la traduzione assistita e automatica

Nel presente capitolo vengono introdotte e trattate le principali questioni teoriche inerenti alle tecnologie per la traduzione. Inizialmente si definiscono concetti e termini chiave per una corretta trattazione del tema e il posizionamento della presente ricerca rispetto alla letteratura di riferimento. Segue una sezione dedicata alla traduzione assistita, in cui vengono esaminati lo stato dell'arte e le principali caratteristiche dei CAT tools. Il capitolo prosegue con un approfondimento sulla traduzione automatica. Si parte da un breve excursus storico sullo sviluppo della traduzione automatica, quindi si forniscono una classificazione e una descrizione dei principali approcci di traduzione automatica concentrandosi sulla traduzione automatica neurale e sulle sue caratteristiche tecniche. Conclude il capitolo un breve accenno sull'intelligenza artificiale.

1.1 Concetti e termini

Nel corso degli anni si è assistito allo sviluppo di diverse tecnologie per la traduzione e il loro impiego da parte di professionisti¹ nel settore della traduzione è aumentato in modo esponenziale. Tali tecnologie possono essere definite come segue: “[...] eine Reihe von Werkzeugen, die von Humanübersetzern, Revisoren, fachlichen Prüfern und anderen Personen zur Erleichterung ihrer Arbeit genutzt werden”² (DIN EN ISO 17100 2015: 24, cit. in Krüger 2019: 145). Si tratta pertanto di strumenti che possono essere implementati nel *workflow* di figure professionali attive nel settore dei servizi linguistici con l'obiettivo di trarre vantaggi e benefici in termini di carico lavorativo e rendimento.

Tra le tecnologie per la traduzione, Monti (2014: 55) annoverano le risorse e gli strumenti per la traduzione automatica (TA) e assistita. La letteratura scientifica fornisce numerose definizioni e teorizzazioni per categorizzare sia i concetti di traduzione automatica e assistita, sia l'interazione tra l'uomo e la macchina, che ne costituisce il fondamento. Hutchins e Somers (1992: 148) illustrano il rapporto tra la macchina e l'essere umano in ambito traduttivo come segue (cfr. Fig. 1):

¹ Per ragioni di scorrevolezza espositiva, nel presente elaborato si ricorre al genere grammaticale maschile con valore inclusivo, intendendolo come comprensivo di tutte le identità di genere. Tale scelta stilistica non intende in alcun modo escludere o discriminare.

² “[...] una serie di strumenti impiegati da traduttori umani, revisori linguistici, revisori specialistici e altre persone per facilitare il proprio lavoro” (trad. mia).

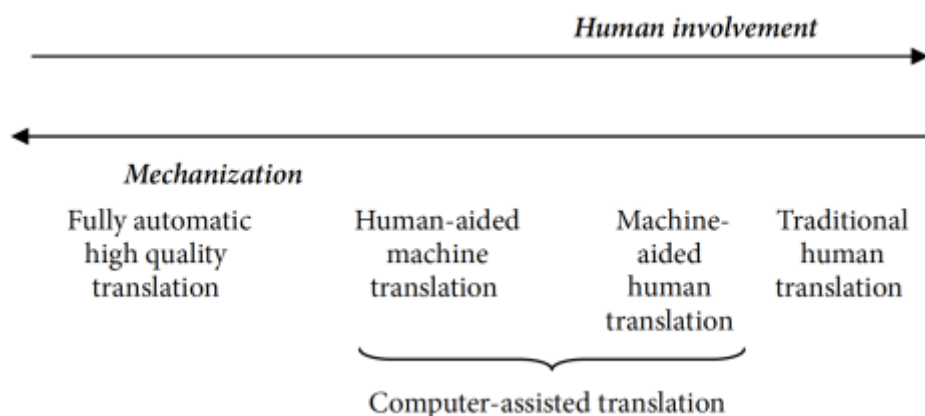


Figura 1: Traduzione umana e traduzione automatica, Hutchins e Somers (1992:148)

Nella classificazione di Hutchins e Somers la traduzione umana e la traduzione automatica sono collocate agli estremi di un continuum che prende in considerazione diverse forme d'interazione uomo-computer. Nella Figura 1 è possibile osservare all'estremità sinistra la denominazione “traduzione di alta qualità completamente automatica” (*fully automatic high quality translation* – FAHQT) e all'estremità destra la denominazione “traduzione umana tradizionale” (*traditional human translation*). Nel mezzo si collocano la traduzione automatica assistita dall'uomo (*human-aided machine translation* – HAMT) e la traduzione umana assistita da computer (*machine-aided human translation* – MAHT). Secondo Hutchins e Somers (1992: 148), questi due procedimenti traduttivi rientrano nell'ampio concetto di “traduzione assistita” (*computer-assisted translation* – CAT).

Diverso invece è l'approccio di Reinke (2004: 73-74), secondo cui nell'ambito della TA si possono individuare i seguenti concetti: la traduzione umana assistita da computer, la traduzione automatica assistita dall'uomo e la traduzione completamente automatica (*fully automated machine translation* - FAMT), nota anche come traduzione di alta qualità completamente automatica o traduzione automatica (*automatic translation*). Analogamente alla FAHQT individuata da Hutchins e Somers (1992), la FAMT di Reinke (2004) viene intesa come una traduzione di elevata qualità eseguita interamente da un software, e priva di qualsiasi intervento umano, quindi di interazione uomo-macchina. Tuttavia, tra i due approcci, quello di Hutchins e Somers e quello di Reinke, emerge una differenza concettuale significativa: mentre Hutchins e Somers considerano la traduzione assistita (traduzione automatica assistita dall'uomo e traduzione umana assistita da computer) come un'area distinta dalla TA, Reinke tende a ricondurre gli approcci di traduzione assistita all'interno della più ampia categoria della TA.

In controtendenza rispetto alla visione di Reinke sopra illustrata, negli ultimi anni la letteratura specialistica ha portato ad associare la TA principalmente alla traduzione completamente

automatica e alla traduzione automatica assistita dall'uomo, ponendo tali approcci in una relazione di netta contrapposizione con gli strumenti di traduzione assistita (cfr. Hutchins 2003b: 161)). In linea con questa posizione espressa da Hutchins (2003b), nel presente studio si adotta un approccio che distingue chiaramente tra TA e CAT, configurandoli come due concetti autonomi e non sovrapponibili, in quanto l'uno non rientra nella definizione formulata per l'altro. Pertanto, il presente studio intende discostarsi dalla tesi secondo cui la traduzione automatica assistita dall'uomo rientri nel concetto di TA.

A tal fine, si prende in esame il termine “traduzione automatica” inteso come “quel processo messo in atto da un software che, dato un testo di partenza, produce automaticamente un testo di arrivo senza nessun intervento umano” (Monti 2019: 15). In breve: l'obiettivo e la sfida della TA è rendere meccanico un processo umano. Questo è quanto emerge anche nella definizione fornita dall'European Association of Machine Translation (EAMT)³ che definisce la TA, corrispondente al termine inglese *Machine Translation*, come segue: “Machine Translation (MT) is the application of computers to the task of translating texts from one natural language to another.”⁴ (EAMT)⁵.

Le due definizioni precedentemente riportate, fornite rispettivamente da Monti (2019: 15) e dall'EAMT, descrivono sistemi di TA che traducono in autonomia. La presente indagine adotta tale accezione ristretta della TA, non intendendo ricondurre la traduzione automatica assistita dall'uomo alla TA, ma considerandola come una sua derivazione.

Per comprendere come viene intesa in questa sede la traduzione automatica assistita dall'uomo, torna utile quanto affermato da Scarpa (2008: 304), che definisce questo approccio come quel procedimento in cui “il traduttore interviene in aiuto al sistema di traduzione automatica prima, durante e dopo il processo di traduzione”. L'intervento del traduttore può dunque avvenire sia durante la fase di *pre-editing*, che in quella di *post-editing*. Con il termine “pre-editing” si intende la fase in cui il traduttore prepara il testo alla TA allo scopo di ottimizzare le prestazioni del sistema di TA utilizzato. Diversamente, il termine “post-editing” definisce la fase di correzione del testo ad opera del traduttore, il quale revisiona l'output cioè il testo tradotto dal sistema di TA apportando le dovute modifiche (Scarpa 2008: 304-305). In alcuni sistemi di TA interattiva, è inoltre previsto un intervento umano anche durante il processo traduttivo vero e proprio, ad esempio per la risoluzione di ambiguità lessicali o strutturali (Scarpa 2008: 305). Pertanto, la traduzione automatica assistita dall'uomo prevede un affiancamento dell'utente al sistema automatico, volto a supportarne il funzionamento.

³ <https://eamt.org/> (ultima consultazione: 20/12/2025)

⁴ La definizione appena citata pone l'accento anche sull'espressione “lingua naturale” (*natural language*) che si contrappone all'espressione “lingua artificiale” (*artificial language*). Pertanto, il sistema di traduzione automatica (di seguito “sistema di TA”) traduce tra due lingue naturali, ovvero impiegate dagli uomini per la comunicazione.

⁵ <https://eamt.org/what-is-machine-translation/> (ultima consultazione: 20/12/2025)

Differente è invece l'approccio della CAT, che, come spiegato precedentemente, si contrappone alla TA e alla traduzione automatica assistita dall'uomo in quanto il termine designa un sistema in cui “the translator [is] provided with software and other computer-based facilities to assist in the task of translation, which [remains] under the control of the human” (Somers 2003: 6). Si tratta, dunque, di un procedimento di traduzione in cui il traduttore è assistito nello svolgimento del suo lavoro da programmi informatici, comunemente chiamati CAT tools.

Alla luce delle definizioni sopra riportate, si può dedurre che la discriminante fondamentale tra la TA e la CAT sia l'attribuzione della responsabilità del processo traduttivo. Nella TA un computer svolge la traduzione, mentre l'intervento umano è eventualmente demandato a una fase successiva di correzione dell'output. Al contrario, nella CAT il traduttore umano svolge la traduzione, avvalendosi di strumenti computerizzati concepiti per supportarlo nello svolgimento della sua attività (Bowker 2002). In sintesi, mentre la TA mira a una sostituzione del traduttore, la CAT si configura come uno strumento al suo servizio (Bowker 2002).

1.2 Traduzione assistita

Il termine “sistema di traduzione assistita” (CAT tool) indica un software che consente di “recuperare porzioni di testo già tradotte ed equivalenze terminologiche interlinguistiche già individuate” (Lecci e Di Bello 2012: 10). In una definizione più ampia, i CAT tools possono essere visti come tutti gli strumenti computerizzati che il traduttore usa come aiuto per lo svolgimento del suo lavoro (Bowker 2002).

Nel corso degli anni, la traduzione assistita e i CAT tools hanno fortemente trasformato la professione del traduttore. Un'indagine condotta da Pielmeier e O'Mara (2020) su un campione di circa 7000 traduttori ha rilevato che l'88% degli intervistati impiega regolarmente memorie di traduzione o CAT tools durante lo svolgimento del proprio lavoro. Tale studio evidenzia il ruolo ormai imprescindibile ricoperto dalla traduzione assistita.

Gli strumenti che hanno “assistito” il traduttore nel suo lavoro si sono evoluti dal 1600 ad oggi permettendo di ottenere risultati di qualità progressivamente più elevata in tempi sempre più ridotti. Tuttavia, la traduzione assistita, come è intesa attualmente, è un concetto di recente sviluppo. Usualmente la sua origine viene individuata agli inizi degli anni '80 del secolo scorso e precisamente quando il computer è stato integrato nel *workflow* del traduttore e sono nati i primi software dedicati all'attività di traduzione (Lecci e Di Bello 2012: 11).

Negli anni Settanta alcune aziende hanno iniziato a lavorare allo sviluppo delle prime banche dati terminologiche. Il primo sistema di cui si ha conoscenza è TSS (Translation Support System), un software sviluppato dalla software house statunitense ALPS. TSS fu realizzato attorno

alla metà degli anni '80 e fu adottato da alcune grandi società, tra le quali IBM, per la propria attività di traduzione interna. Nella seconda metà degli anni '80 la società olandese INK realizzò un sistema ispirato a TSS: Text Tools. Negli stessi anni Star AG sviluppò Transit, un software nato per uso interno. La produzione di questo tipo di strumenti in questi anni fu estremamente prolifera, tuttavia nessuno di questi si diffuse su larga scala a causa degli eccessivi costi. Gli anni '90, invece, si caratterizzano per lo sviluppo di Trados Multiterm avvenuto nel 1990. Nel 1992 venne lanciato Trados Translator's Workbench per DOS e IBM presentò SAA AD/Cycle Translation Manager/2, mentre nel 1993 venne introdotto Dèja Vu e nel 1994 Star (Lecci e Di Bello 2012: 11-12).

Tuttavia, il mercato vide un vero forte sviluppo solo a partire dagli anni 2000. Come sostengono Lecci e Di Bello (2012: 12), “una svolta storica è rappresentata [...] dalla fusione tra SDL e Trados (2005) con la nascita di SDL Trados”. Tuttora la società inglese SDL Trados è leader nel settore e SDL Trados Studio è il CAT tool più utilizzato dai traduttori professionisti⁶.

Di seguito è presentato un approfondimento sulla traduzione assistita e sui suoi strumenti.

1.2.1 CAT tools

Generalmente i CAT tools presentano tre strumenti essenziali: una memoria di traduzione (*Translation Memory*, TM) ossia un “database fraseologico bilingue contenente porzioni di testo in lingua di partenza (SEGMENTI SOURCE) e porzioni di testo in lingua di arrivo (SEGMENTI TARGET)” (Lecci e Di Bello 2012: 9), un database terminologico bilingue o multilingue (*Termbase*, TB), cioè un vero e proprio dizionario costituito da un *source term* in una lingua di partenza e l'equivalente *target term* in una lingua di arrivo, e l'editor di testo, cioè l'interfaccia su cui viene eseguita la traduzione. Le singole componenti saranno approfondite di seguito.

Come indicato da Lecci e Di Bello (2012), è possibile suddividere i CAT tools in due macrocategorie: gratuiti o proprietari/commerciali. I CAT tools gratuiti possono essere scaricati e usati direttamente sul proprio dispositivo senza l'obbligo di acquisto di una licenza d'uso e si dividono in non *open-source* e *open-source*. In quest'ultimo caso, viene messo a disposizione dell'utente il codice sorgente, consentendogli di adattare lo strumento alle proprie esigenze. Un esempio di CAT tool *open-source* è OmegaT. MateCat è invece un esempio di CAT tool gratuito, non *open-source*.

I CAT tools proprietari, definiti anche commerciali, sono strumenti il cui utilizzo è subordinato all'acquisto di una licenza d'uso: per poterli installare e utilizzare è solitamente necessario effettuare un pagamento, che può avvenire sotto forma di licenza perpetua (una tantum)

⁶ Attualmente Trados Studio appartiene a RWS, società britannica che fornisce anche servizi di traduzione. Per ulteriori informazioni, si veda <https://www.trados.com/it/about/> (ultima consultazione: 20/12/2025).

o più frequentemente sotto forma di abbonamento periodico (mensile o annuale). I CAT tools proprietari sono strumenti sviluppati e distribuiti da aziende che ne detengono i diritti esclusivi e, a differenza dei tools *open-source*, il loro codice sorgente non è accessibile né modificabile dall'utente. Tra i CAT tools a pagamento, i più conosciuti sono SDL Trados Studio, MemoQ e Wordfast⁷. Di norma, un traduttore professionista utilizza CAT tools proprietari, che offrono funzionalità più avanzate; tuttavia, alcuni preferiscono strumenti *open-source* per la loro maggiore semplicità d'utilizzo e gestione. Inoltre, quasi tutti i CAT tools assicurano un certo livello di interoperabilità grazie all'uso di formati di scambio universali (Lecci e Di Bello 2012: 12).

Allo stato attuale dell'arte, la gran parte dei CAT tools, sia proprietari sia gratuiti, consentono l'integrazione di sistemi di TA.

- Memoria di traduzione

Per memoria di traduzione (TM) si intende “ [...] a type of linguistic database that is used to store source texts and their translations” (Bowker 2002: 92). La TM è dunque un archivio multilingue che contiene e associa segmenti in una lingua di partenza con le corrispondenti traduzioni in una lingua di arrivo. Le coppie di segmenti formano le unità di traduzione (*Translation Unit*, TU), cioè l'unità di misura della TM (Lecci e Di Bello 2012: 15).

Come afferma Scarpa (2008: 305), la memoria di traduzione opera secondo il principio del “cerca e sostituisci”: durante lo svolgimento della traduzione suggerisce automaticamente proposte di traduzione memorizzate al suo interno quando frasi uguali o simili occorrono nel testo di partenza. La traduzione proposta prende il nome di *translation candidate*. Nel caso in cui la frase da tradurre coincida totalmente con quella ritrovata nella memoria di traduzione, si parla di *100% match*, distinguibile a sua volta in *exact match* o *full match*. Un *exact match* si verifica quando la proposta di traduzione della TM coincide con il segmento di partenza sia dal punto di vista linguistico che da quello della formattazione. Si parla di *full match*, invece, quando il segmento di partenza e il *translation candidate* differiscono unicamente per i cosiddetti *placeble elements*, cioè elementi variabili come numeri, date, valute, nomi propri, ecc. (Bowker 2002).

Qualora la corrispondenza sia solo parziale, si genera un *fuzzy match*. La TM propone pertanto un *translation candidate* che può essere modificato in modo tale da ottenere la traduzione desiderata della frase di partenza. In seguito alla modifica, i segmenti nella lingua d'arrivo e di partenza vengono introdotti all'interno del database come nuova unità di traduzione (Scarpa 2008: 305). Il *fuzzy match* ha solitamente una percentuale minima di corrispondenza del 70% che viene

⁷ Di questi tools sono disponibili versioni di prova gratuite ma limitate nel tempo e/o nel numero di caratteri gestibili. Wordfast offre una versione online (Wordfast Anywhere) con funzionalità gratuite ma limitate.

definita *minimum match value* e costituisce il valore sulla cui base la TM propone o no un *translation candidate* durante il processo di traduzione. Il valore minimo (70%) è stabilito di default, tuttavia può essere modificato dall'utente in base alle sue specifiche esigenze. Comunque, i CAT tools offrono funzionalità avanzate per il recupero dei *matches* anche al di sotto del *minimum match value* (Lecci e Di Bello 2012: 15-16).

Nel caso in cui la TM non formuli alcuna proposta di traduzione, il traduttore dovrà inserire da zero una traduzione e in questo caso si parlerà di *no match* (Lecci e Di Bello 2012: 15).

Sulla base di quanto esposto fino ad ora, è possibile individuare due metodi di impiego della memoria di traduzione. Il primo è il metodo incrementale, cioè il traduttore crea una memoria di traduzione mentre svolge il suo lavoro di traduzione. La TM cresce autonomamente e di pari passo con lo svolgimento della traduzione da parte dell'utente. Pertanto, i segmenti della lingua di partenza e l'equivalente traduzione nella lingua d'arrivo vengono automaticamente inseriti nella TM costituendo le unità di traduzione della stessa. Il secondo metodo consiste nel riciclo di contenuti e prevede il recupero di traduzioni già svolte attraverso applicazioni che consentono di allineare i testi tradotti precedentemente ed estrapolare i contenuti sotto forma di segmenti allineati.

- Database terminologico

Come secondo elemento costitutivo è possibile individuare il database terminologico, anche chiamato in inglese *termbase* (TB). Lecci e Di Bello (2012: 16) definiscono il TB come un database multilingue aggiornabile, contenente termini appartenenti a un dominio specifico in una lingua di partenza ed equivalenti traduttivi in una o più lingue di arrivo. Pertanto, è possibile creare un TB per un dominio specifico e per una o più coppie linguistiche che abbiano la stessa lingua di partenza.

Il database terminologico permette la gestione terminologica che viene intesa come “la preparazione e la gestione di raccolte di termini che permettono al traduttore la ricerca, il riconoscimento e l'inserimento automatico della terminologia appropriata nella propria traduzione” (Scarpa 2008: 311). Per i traduttori professionisti, il tool di gestione terminologica è uno strumento che supporta e facilita lo svolgimento del lavoro garantendo soprattutto coerenza terminologica, che risulta indispensabile in particolar modo nel caso di domini specialistici.

Il TB funziona nel seguente modo: quando il software individua nel testo da tradurre un termine contenuto nel *termbase*, propone al traduttore l'equivalente nella lingua di arrivo (Zanchetta 2011: 47). In questo caso, durante l'esecuzione della traduzione, il TB viene interrogato automaticamente dal CAT tool e restituisce in tempo reale equivalenti interlinguistici utili. Oltre alla ricerca di tipo automatico, è possibile impiegare il TB effettuando una ricerca manuale.

Il database terminologico consente inoltre di organizzare i termini in schede terminologiche e glossari. Le schede terminologiche sono uno strumento più complesso rispetto ai glossari, in quanto non solo permettono di inserire termini nella lingua di partenza e gli equivalenti nella lingua d'arrivo, ma anche ulteriori informazioni come note grammaticali, contesto, immagini esplicative, livello di affidabilità del termine, o altro. Come già indicato, invece, i glossari sono risorse minime costituite da termini equivalenti in una o più lingue ed eventuali note.

- Editor di testo

Il terzo elemento che compone i CAT tools è l'editor di testo. Con questo termine si intende l'interfaccia in cui l'utente esegue la traduzione. Questo spazio permette al traduttore di visualizzare contemporaneamente il segmento da tradurre e la traduzione in corso, le proposte di traduzione della TM e del TB ed eventuali ulteriori strumenti messi a disposizione dal CAT tool che si impiega (Lecci e Di Bello 2012: 17).

Nell'editor di testo, quest'ultimo viene suddiviso in segmenti in virtù di regole basate principalmente sulla punteggiatura. Un'unità di traduzione solitamente corrisponde a una frase, ma può coincidere anche con un titolo, una voce di un elenco o una cella di una tabella. Ne consegue che l'unità di traduzione possa identificarsi, a seconda dei casi, con una frase, un segmento di frase o anche un intero paragrafo (Bowker 2002).

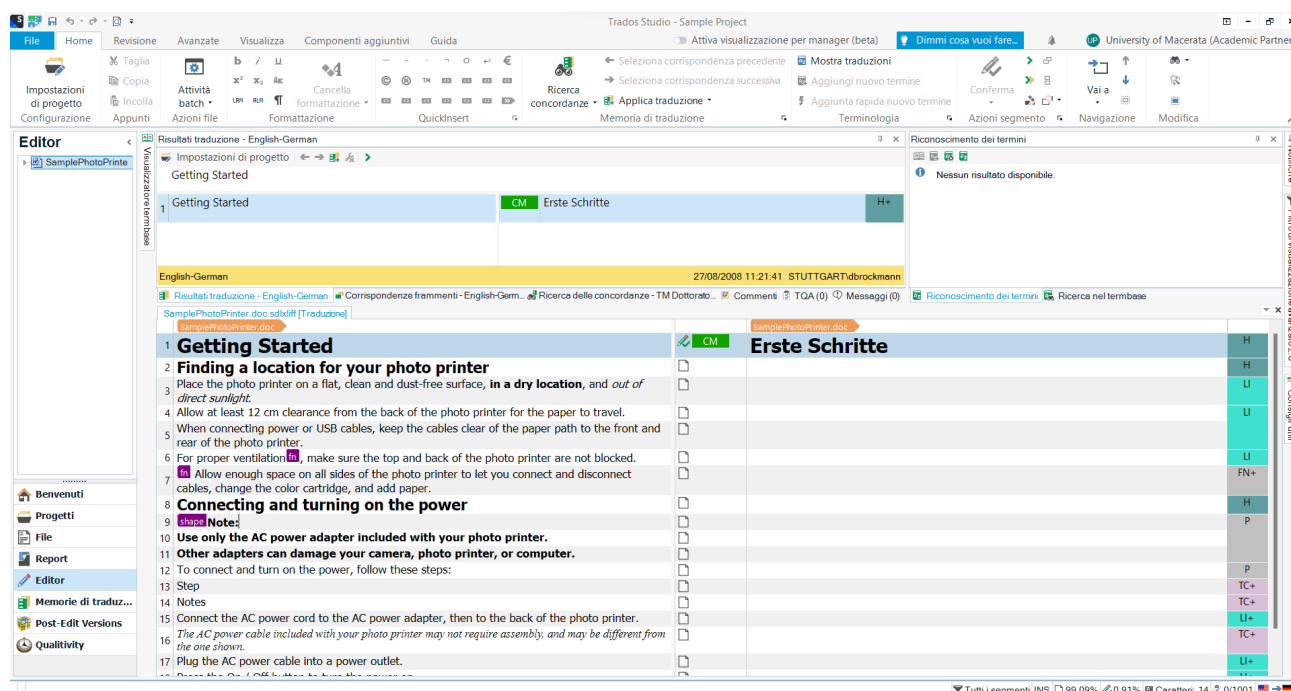


Figura 2: Editor di testo di Trados Studio

La Figura 2 illustra un esempio dell'editor di testo del programma Trados Studio. Il testo da tradurre compare nella parte inferiore, sul lato sinistro dell'interfaccia, ed è diviso nella seconda colonna a sinistra in segmenti che sono stati numerati nella prima colonna a sinistra. Sul lato destro invece l'utente può inserire la sua proposta di traduzione. Le proposte di traduzione della TM e del TB possono essere consultate dall'utente nei riquadri superiori che mostreranno i suggerimenti riguardanti il segmento selezionato. La colonna centrale fornisce informazioni sullo stato della traduzione del segmento: un foglio di carta bianco sta per "segmento non tradotto"; la matita per "bozza"; la matita con una spunta verde per "segmento tradotto e confermato". Accanto a questi simboli appare un riquadro che fornisce informazioni sul *match*: la sigla CM indica il *context match*; la sigla AT sta per traduzione automatica; se il riquadro contiene invece un numero, questo fornisce informazioni sulla percentuale del *match*: 100% per i *full* o *exact match*; da 99% fino al *minimum match value* per i *fuzzy match*. Infine, l'ultima colonna a destra dà informazioni sulla struttura del testo, ad esempio H sta per titolo; P sta per paragrafo e FN sta per nota a piè di pagina.

1.2.2 Vantaggi e svantaggi nell'uso dei CAT tools

Allo stato attuale i CAT tools costituiscono una risorsa indispensabile per i traduttori professionisti, in quanto forniscono una serie di irrinunciabili vantaggi.

In primo luogo, accelerano il processo di traduzione consentendo un guadagno in termini di tempo: come già espresso precedentemente, questi strumenti permettono il recupero di porzioni di testo già tradotte che quindi il traduttore non dovrà ritradurre ma solo controllare. Ciò garantisce anche coerenza a livello terminologico e fraseologico (Lecci e Di Bello 2012: 10-11). Infatti, grazie all'uso della TM, è possibile riutilizzare la stessa traduzione per parti di testo uguali; mentre, come già menzionato, il TB assicura la resa uniforme dei termini già tradotti nelle loro occorrenze successive.

I CAT tools consentono, inoltre, la gestione di progetti di traduzione complessi e supportano e facilitano la gestione di formati strutturati, come ad esempio i file html. In questo caso, il software permette in particolare di modificare solo i contenuti traducibili, l'integrità del codice e della struttura originaria. Inoltre, l'utente può gestire i progetti di traduzione in modo più efficiente, creandoli e dividendoli per dominio o cliente. Non ultimo, si rileva un beneficio economico: consentendo il recupero e il riuso di parole e segmenti già tradotti, si riducono sensibilmente i costi a carico dei clienti (Lecci e Di Bello 2012: 10-11)

Tuttavia, l'uso dei CAT tools presenta anche limiti evidenti. La loro efficacia diminuisce drasticamente in determinate tipologie testuali, in primo luogo in quelle in cui la variazione lessicale e fraseologica risulta un requisito importante, come nella saggistica e nei romanzi.

Analogamente, ad esempio nei testi giuridici, l'eccessiva lunghezza dei periodi può ostacolare la generazione di *matches*. Gli strumenti attualmente disponibili assistono il traduttore nella resa di un'ampia gamma di testi, rimangono tuttavia particolarmente utili per i “testi molti ripetitivi e standardizzati come la documentazione tecnica e di istruzione, specialmente quando si tratta di produrre un aggiornamento di una versione precedente di un prodotto, dove le frasi che si ripetono – identiche o simili – sono molto frequenti.” (Scarpa 2008: 306).

1.3 Traduzione automatica

Il termine “sistema di TA” descrive il software che fornisce la traduzione. Tra sistemi di TA maggiormente noti e impiegati si individuano Google Translate, DeepL e Bing Microsoft Translator. Col termine “input” si fa riferimento al testo di partenza che viene immesso nel sistema di TA. Il termine “output” è utilizzato invece per indicare il testo prodotto dal sistema di TA in seguito alla traduzione dell'input. In maniera più specifica si parla di “output grezzo” per riferirsi al testo di arrivo generato dal sistema di TA, privo modifiche apportate dall'utente per migliorarne la qualità (Gaspari 2011: 14).

Sebbene la diffusione della traduzione automatica tra il grande pubblico sia relativamente recente, il suo sviluppo risale a molto tempo fa. Il punto di partenza della sua evoluzione può essere individuato nell'avvento del calcolatore. Un ruolo essenziale è stato svolto anche dalla diffusione di quelle che Monti (2019: 55) definisce le “idee sugli universali linguistici”. Queste idee costituiscono le regole, comuni a tutte le lingue naturali, utili allo sviluppo di “software in grado di tradurre da una lingua naturale a un'altra senza alcun intervento da parte dell'uomo” (Monti 2019: 55).

Gli storici fanno comunemente risalire l'origine della TA al 1949 e più precisamente negli Stati Uniti con il *Weaver Memorandum*. Nella versione seguente, intitolata *Translation*, Warren Weaver (1949), direttore della divisione di scienze naturali della fondazione Rockefeller, delinea le prospettive della TA. Il documento si compone di 12 pagine e affronta la sperimentazione della traduzione parola-per-parola individuandone metodi e strategie. Nel suo memorandum Weaver riconosce l'esistenza di molteplici lingue come un problema e un ostacolo allo scambio culturale. Partendo da questa convinzione, per risolvere il problema individua uno strumento automatico che possa tradurre grandi quantità di testi tecnici. Weaver si limita a individuare una soluzione che riguardi esclusivamente i testi tecnici poiché riconosce che in questi il problema della polisemia è meno rilevante e i termini impiegati sono solitamente monoreferenziali (Monti 2019: 56).

All'inizio del suo memorandum l'autore dichiara “(it) has been written with one hope only – that it might possibly serve in some small way as a stimulus to someone else, who would have the

techniques, the knowledge, and the imagination to do something about it.” (Weaver 1949: 1). La speranza che Weaver esterna con le sue parole si tramutò presto in una grande fiducia nei confronti della TA. Il documento ebbe grande successo e si iniziò a pensare che presto i traduttori umani sarebbero stati sostituiti dai computer. Questa convinzione si rafforzò ulteriormente nel 1954, quando fu sviluppato il primo sistema di TA grazie alla collaborazione tra IBM e la Georgetown University, aumentando l’aspettativa di un sistema capace di generare traduzioni di elevata qualità in modo del tutto autonomo (Hutchins 2004).

Tuttavia, verso la fine degli anni ‘50 l’ottimismo si dissipò e ci si rese conto che sarebbe stato impossibile raggiungere una TA che avesse una qualità tale da sostituire i traduttori umani (Scarpa 2008: 302). A tal riguardo Bar-Hillel (1960: 117) affermò che la TA di alta qualità sarebbe stata irraggiungibile. Questa consapevolezza determinò una significativa riduzione dell’interesse nei confronti della TA. Una prova di ciò fu il Rapporto ALPAC, *Automatic Language Processing Advisory Committee*, realizzato nel 1966 in cui si prendono in esame la reale fattibilità e le concrete possibilità di successo della TA tenendo conto di alcuni aspetti, e più precisamente dei costi, della qualità e delle prospettive future. Il risultato stilato nel rapporto è un giudizio severo nei confronti della TA, che viene definita come costosa, lenta e inutile perché incapace di imitare il modo di lavorare del traduttore umano (Monti 2019: 61). Simbolica è la seguente affermazione pronunciata da un traduttore e riportata nel Rapporto:

I must confess that the results were most unhappy. I found that I spent at least as much time in editing as if I had carried out the entire translation from the start. Even at that, I doubt if the edited translation reads as smoothly as one which I would have started from scratch. I drew the conclusion that the machine today translates from a foreign language to a form of broken English somewhat comparable to pidgin English. But it then remains for the reader to learn this patois in order to understand what the Russian actually wrote. Learning Russian would not be much more difficult. (Rapporto ALPAC 1966)

Il rapporto ebbe effetti devastanti: nell’opinione pubblica si diffuse un’ampia sfiducia nei confronti della TA e questo negli Stati Uniti ebbe come conseguenza un drastico taglio dei fondi di ricerca nel settore per circa 20 anni (Hutchins 2003a). Tuttavia, non tutte le comunità scientifiche condivisero il verdetto negativo del rapporto ALPAC. A partire dagli anni ‘70 infatti si poté osservare la nascita di diversi gruppi di ricerca, soprattutto in Canada e in Europa, dove il multilinguismo rappresentava una problematica rilevante (Monti 2019: 61). Un esempio è il progetto TAUM (*Traduction Automatique de l’Université de Montréal*) che portò allo sviluppo del sistema METEO impiegato con successo a partire dal 1977 e che gestiva la traduzione di bollettini

meteorologici (Thouin 1981). Negli stessi anni in Europa anche la Comunità Europea iniziò a supportare diversi progetti, prima adottando il sistema SYSTRAN⁸ e successivamente finanziando un proprio progetto. Nonostante il loro fallimento, furono proprio questi tentativi a dare impulso alla ricerca nel settore della TA in Europa (Monti 2019: 62). Inoltre, come già spiegato ampiamente nel paragrafo 1.2.1, negli anni '80 l'attenzione si concentrò sulla CAT consentendo lo sviluppo dei primi sistemi di memoria di traduzione.

Gli anni '90 segnarono una svolta nel campo della TA, grazie al rinnovato interesse suscitato principalmente dalla diffusione di Internet (Scarpa 2008: 302). Da quel momento il campo della TA fu caratterizzato da un approccio più pragmatico e maggiormente orientato ai bisogni degli utenti. La conseguenza di questa nuova tendenza fu, da un lato, l'abbandono definitivo del mito di una macchina in grado di tradurre come un essere umano e, dall'altro, lo sviluppo di strumenti a supporto del processo di traduzione. A metà degli anni '90 si registrò infatti un notevole aumento del numero delle tecnologie: sistemi di TA, strumenti di traduzione assistita, memorie di traduzione, sistemi online (Monti 2019: 65). L'illustrazione delle diverse tipologie di architettura, a cui sono dedicati i seguenti paragrafi, consentirà di ottenere una panoramica dell'evoluzione della TA. Sarà discusso inoltre come il concetto stesso di TA stia cambiando in risposta alle innovazioni tecnologiche e alle nuove esigenze lavorative dei traduttori professionisti.

1.3.1 Principali architetture dei sistemi di TA

I principali approcci e modelli di TA possono essere suddivisi in due categorie: approcci linguistici e approcci empirici. La traduzione automatica basata su regole e le relative strategie appartengono agli approcci linguistici. La traduzione automatica statistica, la traduzione automatica basata su esempi e la traduzione automatica neurale appartengono agli approcci empirici. Esistono inoltre sistemi ibridi che integrano approcci linguistici e statistici traendo vantaggi da entrambi i metodi (cfr. Monti 2019: 96-99).

1.3.1.1 Dagli approcci linguistici agli approcci statistici

Dai suoi inizi fino agli anni '90, la TA è stata caratterizzata da un approccio denominato traduzione automatica basata su regole (*Rule-Based Machine Translation* – RBMT). Questo approccio fornisce una traduzione, e quindi un testo di arrivo, eseguendo un'analisi del testo di partenza sulla base di risorse linguistiche, dizionari e grammatiche.

⁸ SYSTRAN (da SYStem e TRANslation) è uno dei primi sistemi commerciali di TA. Venne sviluppato nel 1968 negli Stati Uniti nell'ambito di una collaborazione con l'US Air Force, che portò allo sviluppo del primo software di TA dal russo all'inglese. SYSTRAN è tuttora operativo. Per maggiori informazioni si veda: <https://www.systransoft.com/systran/> (ultima consultazione: 20/12/2025).

Secondo Monti (2019: 74-87), si possono distinguere tre differenti metodi di TA basata su regole:

1. *Metodo a traduzione diretta*: sviluppatosi tra gli anni '50 e gli anni '60, si basa su una traduzione diretta tra il testo di partenza e il testo di arrivo senza alcun passaggio intermedio necessario. Servendosi di dizionari bilingue e di un'analisi sintattica elementare, il sistema di TA esegue una traduzione parola-per-parola;
2. *Metodo a interlingua*: si serve di rappresentazioni astratte per generare un testo di arrivo. Tali rappresentazioni hanno la caratteristica di essere comuni a più lingue. L'interlingua è la rappresentazione intermedia che deve contenere le informazioni necessarie per generare un testo indipendentemente dalla lingua di arrivo e dalla lingua di partenza, in quanto essa ha l'aspirazione di essere universale per tutte le combinazioni linguistiche e quindi neutrale rispetto alle lingue di partenza e a quelle di arrivo;
3. *Metodo a transfer*: ha caratterizzato gli anni '80 e si basa su dizionari e grammatiche monolingui e bilingui. Per mezzo di rappresentazioni intermedie astratte dei livelli lessicale, grammaticale e semantico il testo viene elaborato e convertito in un testo di arrivo.

La TA basata su regole rivela limiti evidenti dovuti alla necessità di creare regole lessicali, morfo-sintattiche e semantiche per ogni direzione traduttiva di ogni combinazione linguistica. Si tratta quindi un processo lungo e dispendioso che necessita di personale altamente formato (Gaspari 2011: 24-25).

Verso la fine degli anni '80 la TA basata su regole viene sostituita da approcci di TA chiamati *corpus-based*. Con l'evolversi della tecnologia, le dimensioni di questi corpora cambiano e si inizia a parlare di *data-driven*. In questo approccio rientrano la traduzione automatica basata su esempi (*Example-Based Machine Translation* – EBMT) e la traduzione automatica statistica (*Statistical Machine Translation* – SMT) (Rothwell et al. 2023: 99).

La TA basata su esempi si sviluppa a partire dal 1981. Nel dettaglio, la sua origine può essere ricercata in un articolo di Nagao (1984). L'autore introduce il principio di traduzione per analogia (*analogy translation principle*), riassunto nella seguente citazione:

Man does not translate a simple sentence by doing deep linguistic analysis, rather, man does translation, first, by properly decomposing an input sentence into certain fragmental phrases, (very often, into case frame units) then by translating these phrases into other language phrases, and finally by properly composing these fragmental translations into one long sentence. The translation of each fragmental phrase will be done by the analogy translation principle with proper examples as its reference. (Nagao 1984: 178-179)

Nella citazione sopra riportata si possono individuare i tre elementi costituenti, che Somers (1999: 117) ha successivamente definito come le “tre fasi della TA basata su esempi”:

1. *Matching*: il testo viene analizzato. Le unità di traduzione che compongono il corpus bilingue, cioè il database, fungono da base per l’analisi del testo con l’obiettivo di identificare i segmenti del testo di partenza che sono simili ai segmenti del corpus;
2. *Allineamento*: le unità di traduzione identificate nel testo di partenza e le unità di traduzione identificate nel database bilingue, cioè gli esempi di traduzione, vengono allineati tra di loro;
3. *Ricombinazione*: il testo di arrivo è composto dalle unità di traduzione identificate nel database, che vengono ricombinate tenendo conto della struttura della frase di arrivo.

La TA basata su esempi si basa quindi sulla traduzione per analogia, cioè sull’idea che un testo possa essere diviso in segmenti e che questi vengano tradotti sulla base di esempi tratti da un database. Quest’ultimo è costituito da corpora paralleli, cioè testi originali abbinati alle corrispondenti traduzioni (Monti 2019: 88-89).

Agli inizi degli anni ’90 si è sviluppato l’approccio della TA statistica, rivelatosi da subito come estremamente innovativo; esso ha caratterizzato e dominato la TA fino a qualche anno fa. I sistemi di questa tipologia non impiegano dizionari o grammatiche per la traduzione del testo: tale approccio si basa piuttosto su corpora monolingui e, come precisa Gaspari (2011: 25), su grandi corpora bilingui paralleli. I testi sono allineati tra loro a livello di singole frasi: a ogni frase in una lingua corrisponde la rispettiva traduzione in un’altra lingua.

Addestrato mediante corpora paralleli e bilingui (*training text data*), il sistema individua la traduzione più frequente nella lingua di arrivo per una parola o una stringa di parole contigue della lingua di partenza. I corpora monolingui sono invece impiegati per valutare la probabilità che una determinata parola o stringa di parole corrisponda a un uso effettivo nella lingua di destinazione (Monti 2019: 89-90).

Si può affermare pertanto che l’output della TA statistica sia il risultato di stime quantitative e considerazioni probabilistiche. Il sistema segmenta innanzitutto la frase del testo di partenza in sequenze di parole (*n-grams*). Le ipotesi di traduzione vengono formulate per le sequenze di parole identificate sulla base di un modello di traduzione (*translation model*). Il principio che porta alla formulazione delle ipotesi di traduzione è quello della maggiore frequenza nella lingua di arrivo. La frase nella lingua di arrivo sarà, dunque, il risultato di una ricombinazione dei segmenti tradotti che ricorrono più frequentemente nella lingua stessa (Monti 2019: 89-90).

Originariamente, la TA statistica utilizzava le singole parole come unità di traduzione (Brown et al. 1988). Un’importante evoluzione è stata compiuta in seguito all’impiego di un nuovo metodo

basato su sequenze di parole, che ha fornito da subito risultati migliori (Koehn et al. 2003). Lo sviluppo di quest'ultimo modello ha ispirato molti ricercatori, tra cui Yamada e Knight (2001), contribuendo alla successiva elaborazione del modello sintattico. Sulla base di queste considerazioni, si possono individuare tre diversi modelli di TA statistica:

1. *Modello basato sulla parola (word-based SMT)*: le parole costituiscono le unità di traduzione; di conseguenza, il sistema di TA genera un output sulla base delle corrispondenze lessicali tra il testo di partenza e il testo di arrivo;
2. *Modello basato sulla sequenza di parole (phrase-based SMT)*: le sequenze di parole (*n-grams*) costituiscono le unità di traduzione; il testo di partenza, viene segmentato in sequenze di *n-grams* e tradotto selezionando, all'interno di un corpus di riferimento, la corrispondenza più probabile per ciascuna sequenza;
3. *Modello sintattico (syntax-based SMT)*: il modello sintattico di Yamada e Knight (2001) si serve di *parser*. Il *parser* è un programma informatico che esegue un'analisi automatica della struttura morfologica delle parole sulla base della grammatica e del vocabolario di una determinata lingua. Attraverso quest'analisi, è possibile riconoscere la struttura sintattica della frase e quindi scegliere un ordine delle parole differente per la generazione di una traduzione migliore.

Come evidenziato da Gaspari (2011: 26-27), anche la TA statistica ha dei limiti, legati principalmente ai corpora paralleli usati per l'addestramento, i *training text data*, che appartengono solitamente a tipologie testuali o settori facilmente reperibili e disponibili in più lingue. Pertanto, i sistemi statistici potrebbero restituire risultati non adeguati nel caso in cui questi fossero impiegati per la traduzione di testi appartenenti a tipologie testuali, settori o argomenti diversi da quelli dei dati usati per l'addestramento.

Si è tentato di trovare una soluzione a questi limiti sviluppando sistemi ibridi che uniscono l'approccio linguistico con l'approccio statistico, sfruttando i benefici di un approccio per superare i limiti dell'altro (Gaspari 2011: 26-27; Monti 2019: 96).

Nell'evoluzione dalla TA basata su regole alla TA basata su esempi e alla TA statistica, si passa da una *pull technology* a una *push technology*. Come spiegano Franklin e Zdonik (1998: 516), la *push technology* nasce dall'idea secondo cui “invece di richiedere esplicitamente agli utenti le informazioni di cui hanno bisogno, i dati possono essere inviati agli utenti senza che questi li richiedano espressamente” (trad. mia). Pertanto, la *push technology* applicata alla TA viene intesa come un sistema computerizzato che “spinge” la traduzione verso l'utente, il quale riceve così una soluzione “completa”. È possibile individuare quindi la *push technology* nella TA statistica e nella

TA basata su esempi. L'approccio di tipo *pull technology* caratterizza invece i primi sistemi di TA basata su regole in cui è l'utente a richiedere informazioni specifiche da archivi e collezioni di dati.

In sintesi, la TA statistica ha rappresentato la principale architettura fino all'avvento della traduzione automatica neurale, che presenta differenti punti di forza illustrati nel paragrafo successivo.

1.3.1.2 Traduzione automatica neurale

La traduzione automatica neurale (*Neural Machine Translation* - NMT) rappresenta la tecnologia più recente sviluppata nell'ambito della TA e prevede sistemi basati su reti neurali. Le reti neurali artificiali sono una tecnologia informatica che si ispira e imita le reti neurali biologiche del sistema nervoso e del cervello umano.

Più nello specifico, tali sistemi si basano sulle tecnologie di apprendimento profondo (dall'inglese *deep learning*) e quindi sulle tecnologie dell'Intelligenza Artificiale (IA). Il *deep learning* è una sottocategoria del *machine learning* il cui scopo è realizzare, mediante l'uso di algoritmi, modelli di reti neurali capaci di replicare l'apprendimento umano (Kelleher 2019: 8). Questa caratteristica è la prima delle differenze che intercorrono tra i sistemi di TA neurale e i sistemi di TA statistica. Una seconda differenza può essere individuata nella capacità del sistema di TA neurale di considerare il contesto dell'intera frase. In questo modo, la macchina può ottenere le informazioni sintattiche e semantiche necessarie per generare la traduzione (Monti 2019: 102-103). Un aspetto in comune che invece caratterizza sia la TA statistica che la TA neurale è che entrambe rientrano nella tipologia di *push technology*.

Il rilascio del primo sistema facente parte di questa tipologia è avvenuto nel 2016 da parte di Google Brain, il gruppo di ricerca sulle applicazioni di intelligenza artificiale. Google è stato seguito da Systran e successivamente da Microsoft.

Le strutture delle reti neurali artificiali sono costituite da gruppi di neuroni artificiali collegati tra loro che formano il cosiddetto "strato" (dall'inglese *layer*). I neuroni di uno strato ricevono informazioni dai dati di addestramento e le trasmettono ai neuroni dello strato successivo fino al raggiungimento del risultato finale, l'output (cfr. Fig. 3). La Figura 3, riportata di seguito, illustra i diversi livelli della rete neurale, divisi da Monti (2019: 100) in tre distinte categorie:

1. *Input layer* (strato di ingresso): rappresenta il primo strato della rete neurale;
2. *Hidden layer* (strato nascosto): rappresenta lo strato intermedio tra lo strato di ingresso e lo strato di uscita. La rete si compone potenzialmente di più strati nascosti che ricoprono un ruolo essenziale per il *deep learning*. Questi strati determinano la complessità della rete:

maggiore è il numero di strati, maggiore è la complessità della rete stessa. In questi livelli si eseguono le procedure di calcolo utili alla generazione della traduzione.

3. *Output layer* (strato di uscita): rappresenta l'ultimo strato della rete che fornisce il risultato.

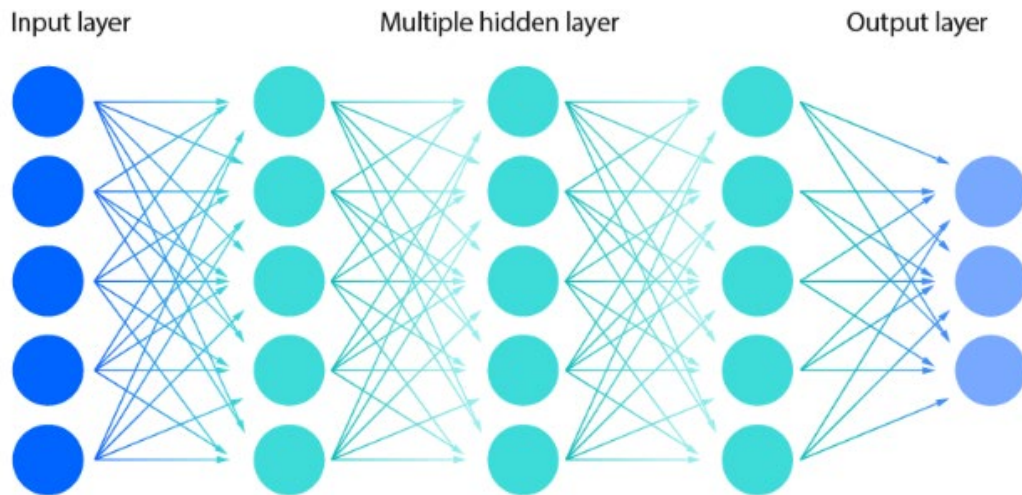


Figura 3: Illustrazione di una rete neurale, tratto da <https://www.ibm.com/it-it/think/topics/neural-networks>
(ultima consultazione: 20/12/2025)

Per comprendere al meglio il funzionamento delle reti neurali artificiali, si può pensare ai neuroni come a nodi della rete. Ogni nodo è collegato ai nodi dei livelli successivi. Ogni collegamento della rete riceve l'input e lo moltiplica per un valore che corrisponde all'effetto che un determinato nodo ha su un altro nodo. L'informazione è trasmessa per mezzo di stimoli che inibiscono o attivano i neuroni dei diversi strati.

I neuroni artificiali hanno un valore numerico. La seguente formula può essere utilizzata per calcolare il valore numerico di un neurone artificiale collegato ad altri neuroni (numerati 1,2,3, ecc.) (Forcada 2017: 293):

$$Y = w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3 + \dots + w_n \times x_n + b$$

Il valore x rappresenta lo stato del neurone, cioè l'attivazione del neurone collegato a quello di cui si vuole calcolare il valore numerico; il valore b (dall'inglese *bias*) rappresenta la tendenza del neurone ad attivarsi; il valore w (dall'inglese *weight*) indica il peso del collegamento neurale. Il peso può essere positivo o negativo: se lo stimolo viene ricevuto attraverso una connessione con un peso positivo, il neurone stimolato tende ad attivare quello ad esso collegato; se lo stimolo viene ricevuto attraverso una connessione con un peso negativo, il neurone stimolato tende a inibire quello ad esso collegato (Forcada 2017: 293).

Una rete neurale può essere *fully connected* (completamente connessa) o *partially connected* (parzialmente connessa). Si parla di una rete completamente connessa quando ogni singolo nodo di uno strato della rete è collegato con ogni singolo nodo dello strato successivo; se invece manca anche un solo collegamento tra il nodo di uno strato e quello di uno strato successivo, si parla di rete parzialmente connessa (Haykin 2009: 23).

1.3.1.2.1 *Word embedding*

Prima di affrontare in modo più approfondito le architetture delle reti neurali, è necessario discutere della rappresentazione delle parole all'interno delle reti stesse. Questa trattazione intende facilitare la comprensione del funzionamento delle architetture di reti neurali che saranno illustrate in seguito.

Nel campo dell'elaborazione del linguaggio naturale (NLP), le rappresentazioni delle parole all'interno della rete neurali sono denominate *word embeddings* (Mikolov et al. 2013). Ogni parola viene trasformata in un vettore di numeri reali (solitamente centinaia o migliaia), rappresentato come una lista di valori separati da virgole tra parentesi (es. [1.24, 2.56, 5.23]) (Pérez-Ortiz et al. 2022: 151). Ciascun numero presente nella sequenza di numeri che forma il vettore rappresenta una dimensione. L'idea fondamentale alla base del funzionamento delle reti neurali è di trasformare parole in vettori affinché la geometria dello spazio vettoriale, in cui questi si trovano, rifletta le relazioni semantiche tra parole. Parole con significati simili o che co-occorrono in contesti simili assumono la forma di *embeddings* simili (Pérez-Ortiz et al. 2022: 150-151): nello spazio vettoriale queste parole saranno posizionate più vicine rispetto ad altre parole. In aggiunta a quanto detto, è necessario specificare che più numeri, cioè dimensioni, ha il vettore, più il modello sarà in grado di organizzare le parole secondo vincoli semantici e sintattici complessi (Pérez-Ortiz et al. 2022: 150-151).

La rappresentazione delle parole come *word embedding* consente di cogliere regolarità linguistiche anche tramite alcune semplici operazioni aritmetiche come addizioni e sottrazioni: ad esempio, [re] - [uomo] + [donna] $\simeq$ [regina], oppure [Madrid] - [Spagna] + [Germania] $\simeq$ [Berlino]. L'addizione (o la sottrazione) di due *embeddings* consiste semplicemente nel sommare (o sottrarre) le componenti dei vettori una per una; ad esempio, [1.24, 2.56, 5.23] + [0.12, 1.12, 0.01] = [1.36, 3.68, 5.24] (Mikolov et al. 2013; Pérez-Ortiz et al. 2022: 152).

Nei sistemi di TA neurale i *word embeddings* costituiscono la rappresentazione iniziale delle parole in input. La frase da tradurre viene quindi immessa nelle reti e viene rappresentata come una sequenza di *word embeddings*, che vengono elaborati da una rete neurale per creare una rappresentazione semantica profonda. A partire da questa rappresentazione, il decoder genera la frase nella lingua target (cfr. 1.3.1.2.2). Inoltre, i *word embeddings* vengono appresi

automaticamente durante l'addestramento del modello, insieme ai pesi della rete neurale (cfr. 1.3.1.2.3) (Pérez-Ortiz et al. 2022: 150-151).

Oltre ai *word embeddings* descritti, le reti neurali possono calcolare dei *contextual word embeddings* grazie al meccanismo di attenzione. Questo consente di superare il problema della polisemia delle parole (Pérez-Ortiz et al. 2022: 153-155). Per ottenere un *contextual word embedding* è necessario prima di tutto ottenere dei *non-contextual word embeddings*, uno per ogni parola che forma la frase. Se quindi si prende in esame una frase composta da dieci parole, si otterranno dieci *non-contextual word embeddings*. Successivamente la rete neurale produrrà un *attention vector* che presenterà dieci percentuali indicanti il grado di attenzione da attribuire a ogni parola nella frase per ottenere la rappresentazione della parola polisemica in questione. La posizione delle percentuali nel vettore rispecchia la posizione delle parole nella frase. Un *attention vector* può apparire ad esempio così [11%, 8%, 6%, 25%, 7%, 14%, 9%, 5%, 12%, 3%]. Le percentuali in questo caso sono casuali ma la loro somma deve dare come risultato in ogni caso 100%. Quindi, per ogni parola presente nella frase, sarà calcolato un *attention vector*, che sarà applicato al *non-contextual word embedding* originale per ottenere infine un insieme di dieci nuovi *embeddings*, uno per ogni parola nella frase. Questi costituiscono i *contextual embeddings* in quanto sono influenzati, in misura diversa, dalle ulteriori parole della frase (Pérez-Ortiz et al. 2022: 153-155).

Alla luce di quanto illustrato, si può affermare che le differenze tra i *contextual word embeddings* dipendono dal punto e dall'ordine in cui le parole appaiono nella frase.

1.3.1.2.2 Architetture delle reti

Nel presente paragrafo verranno illustrate le principali architetture di rete neurale, suddivise nelle seguenti categorie:

1. Reti neurali *feed-forward* a singolo strato
2. Reti neurali *feed-forward* multistrato
3. Reti neurali ricorrenti
4. *Transformers*

Una delle caratteristiche dei primi due tipi di architetture elencate, come già suggeriscono le denominazioni, è la presenza di strati. Nel caso della prima architettura, si possono individuare unicamente un *input layer* (strato di ingresso) e un *output layer* (strato di uscita). Le reti neurali *feed-forward* multistrato invece si caratterizzano per la presenza di uno o più *hidden layer* (Haykin 2009: 21-22). Questi modelli sono chiamati *feed-forward* perché le informazioni passano in un'unica direzione, dallo strato di ingresso allo strato di uscita, pertanto non si verificano cicli di feedback in cui gli output del modello vengono reimmessi nello stesso strato (Goodfellow et al.

2016: 164). Tale processo di informazioni in un'unica direzione viene chiamato *forward pass* (Graves 2012: 13).

Dopo le reti neurali *feed-forward*, si sono affermate le reti neurali ricorrenti e i *Transformers*. Una rete neurale ricorrente (RNN) si distingue da una rete *feed-forward* perché ha almeno un ciclo (anche chiamato *loop*) di feedback (Haykin 2009: 23) (cfr. Figura 4). Le connessioni ricorrenti che si creano in una RNN permettono di mantenere una “memoria” degli input precedenti, che possono persistere nella rete e influenzare l'output (Graves 2012: 19). Questo rende le RNN particolarmente adatte per compiti in cui l'ordine dei dati è importante, come:

- Elaborazione del linguaggio naturale (NLP),
- Traduzione automatica,
- Previsione di serie temporali,
- Riconoscimento vocale.

Come sostengono infatti Goodfellow et al. (2016: 367), le RNN rappresentano un'architettura di reti neurali utilizzata per processare dati sequenziali (cfr. Figura 4).

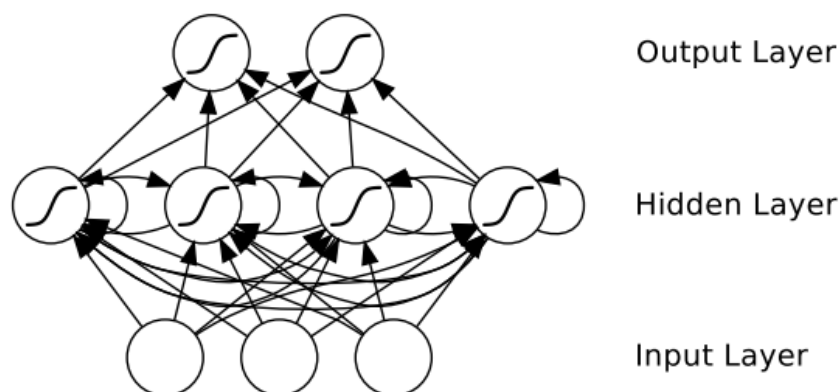


Figura 4: Rete neurale ricorrente, Graves (2012:18)

Alla luce di quanto affermato, una rete neurale *feed-forward* può creare corrispondenze solo tra vettori di ingresso e vettori di uscita, mentre una RNN può in linea di principio creare una mappatura dall'intera storia degli input precedenti a ogni output. L'output di una rete neurale *feed-forward* dipende quindi unicamente dall'input attuale e non da input passati o futuri, come nel caso di un output di RNN (Graves 2012: 13-18). Per questo motivo, le RNN possiedono una “memoria” degli input precedenti.

Queste differenze nel funzionamento hanno impatti rilevanti nelle fasi di addestramento delle reti. Le RNN costituiscono in particolare l'architettura *encoder-decoder*, la quale è composta

da due RNN: una si chiama *encoder* perché codifica una sequenza di simboli (la frase di partenza) in un vettore di lunghezza fissa; la seconda RNN si chiama *decoder* poiché decodifica il vettore in un'altra sequenza di simboli e produce quindi una traduzione (Cho, van Merriënboer, et al. 2014). L'intero sistema *encoder-decoder* viene addestrato congiuntamente per massimizzare la probabilità di una traduzione corretta data una frase di partenza.

Un limite di questa architettura è tuttavia la necessità di comprimere tutte le informazioni della frase in un singolo vettore: ciò compromette le prestazioni delle reti neurali su frasi più lunghe di quelle presenti nel corpus di addestramento, come dimostrato dallo studio di Cho, van Merriënboer, Bahdanau, et al. (2014). Per risolvere questo problema, Bahdanau et al. (2016: 1-2) hanno proposto un'estensione del modello *encoder-decoder* che integra un "meccanismo di attenzione". Questo consente al modello di concentrarsi sulle parti rilevanti della sequenza di input a ogni passaggio di generazione di una parola in una traduzione, evitando di codificare un'intera frase di partenza in un unico vettore di lunghezza fissa. L'*encoder*, quindi, codifica l'input in una sequenza di vettori e seleziona in modo adattivo un sottoinsieme di questi vettori rilevanti per generare ciascuna parola in uscita durante la decodifica della traduzione. Questo meccanismo migliora sensibilmente le prestazioni, soprattutto con frasi lunghe (Bahdanau et al. 2016: 1-2).

Inoltre, il meccanismo di attenzione nelle RNN consente di integrare informazioni dalle parole in ingresso e da quelle già generate; infatti, nelle RNN gli *embedding* contestuali sono calcolati localmente attraverso una doppia direzione di lettura della sequenza (da sinistra a destra e viceversa), come descritto da Pérez-Ortiz et al. (2022: 158-159). Nel caso, ad esempio, di una frase di partenza che si compone di otto parole, gli *embeddings* per la quinta parola codificata, si baseranno sugli *embeddings* delle prime quattro parole, da un lato, e sugli *embeddings* delle tre parole successive, dall'altro. L'introduzione dei moduli di attenzione, originariamente pensati per catturare le dipendenze a lungo termine (Bahdanau et al. 2016; Graves et al. 2014), ha segnato un'evoluzione cruciale rispetto al modello *encoder-decoder* tradizionale.

In questo contesto si inserisce il modello *Transformer*, che rappresenta un punto di svolta per le reti neurali. Si tratta di un'architettura che si differenzia dalle RNN in quanto evita la ricorrenza affidandosi completamente al meccanismo di attenzione per tracciare le dipendenze tra input e output (Vaswani et al. 2023). Il *Transformer* si compone di due moduli: l'*encoder*, che calcola gli *embeddings* contestuali per ogni parola della frase di partenza e il *decoder*, che predice successivamente ciascuna parola della frase di arrivo integrando le informazioni delle parole in ingresso e delle parole di arrivo già generate (Pérez-Ortiz et al. 2022: 156-159). È quindi un modello *encoder-decoder* che in particolare utilizza sia per l'*encoder* che per il *decoder* una struttura composta da strati di *self-attention* impilati e connessi (Voita et al. 2019).

Gli *embeddings* iniziali, ottenuti a partire dai *tokens* d'ingresso e arricchiti tramite informazioni sulla loro posizione nella sequenza, vengono forniti all'*encoder*, strutturato in numero N di strati, ciascuno dei quali contiene due sottolivelli: (a) un meccanismo di *self-attention multi-head*, e (b) una rete *feed-forward* a due strati. Il meccanismo di *self-attention* determina i pesi di attenzione: per ogni parola calcola una nuova rappresentazione basata su una distribuzione di attenzione rispetto a tutte le altre parole della sequenza. Nel meccanismo di attenzione *multi-head*, questo processo viene ripetuto un certo numero di volte con rappresentazioni diverse e il risultato viene concatenato (Voita et al. 2018: 1265-1266). Pertanto, la tecnica di attenzione *multi-head* permette di prestare attenzione congiuntamente a informazioni provenienti da diversi sottostrati di rappresentazione in posizioni diverse.

Analogamente, ogni strato del *decoder* contiene i due sottolivelli sopra menzionati e un ulteriore sottolivello di attenzione *multi-head* che riceve in input l'output dell'*encoder*. Il *Transformer* utilizza quindi il meccanismo di attenzione *multi-head* in tre modalità: strato di *self-attention* dell'*encoder*, strato di *self-attention* del *decoder* e strato di attenzione tra *decoder* ed *encoder* (Voita et al. 2019).

Nel *decoder* l'attenzione viene mascherata per impedire che le posizioni future nella sequenza vengano considerate durante la generazione di ogni parola. Questo meccanismo blocca il flusso "illegale" di informazioni dalle parole successive a quelle correnti, garantendo che ogni parola venga predetta solo in base a quelle già generate (Voita et al. 2018: 1265-1266). In questo contesto, il termine "illegale" si riferisce a un flusso informativo che va da destra verso sinistra, violando, quindi, l'ordine sequenziale proprio dei modelli auto-regressivi: come affermano Vaswani et al. (2023: 5), tramite questo meccanismo si preserva la natura auto-regressiva del modello, cioè la capacità di generare l'output un elemento alla volta, senza accedere ai *tokens* futuri che non sono ancora stati prodotti.

Durante l'addestramento, viene impiegato un corpus parallelo per apprendere pesi, *embeddings* e vettori di attenzione in grado di riprodurre i dati e generalizzare oltre le frasi impiegate durante l'addestramento⁹ (Pérez-Ortiz et al. 2022: 156).

Secondo Pérez-Ortiz et al. (2022: 158) il *Transformer* rappresenta attualmente l'architettura maggiormente impiegata nei sistemi di TA neurali commerciali, in quanto, per i compiti di

⁹ La generalizzazione consiste nella capacità del modello neurale di produrre output corretti anche per frasi mai viste prima, purché simili a quelle usate durante la fase di *training*. Ciò è possibile grazie a una caratteristica delle reti neurali, che garantisce che piccole variazioni negli input producano variazioni limitate negli output. Per ottenere la generalizzazione, è quindi fondamentale che frasi simili abbiano rappresentazioni, cioè *embeddings*, simili e che quindi parole simili siano rappresentate con numeri simili (ossia con *embeddings* vicini tra loro) (Pérez-Ortiz et al. 2022: 149-152).

traduzione, l'addestramento dei *Transformers* è significativamente più veloce rispetto a quello delle RNN (Vaswani et al. 2023: 10).

1.3.1.2.3 *Training*

I sistemi di traduzione automatica neurale (NMT) sono sottoposti a fasi di addestramento durante le quali, tramite l'uso di corpora paralleli e dei modelli linguistici in essi ritrovati, sviluppano le loro capacità. Il *training* richiede una grande quantità di esempi che devono essere presentati più volte e questo lo rende potenzialmente un processo molto complicato.

Secondo Pérez-Ortiz et al. (2022: 148) il *training* è una procedura mediante la quale una rete neurale determina il peso delle connessioni tra i (suoi) neuroni, in modo da generare un output che si avvicini il più possibile agli esempi input-output forniti durante la fase di addestramento.

L'obiettivo, quindi, è ridurre la distanza tra il *gold standard*, cioè le traduzioni realizzate da traduttori professionisti, e l'output generato dal sistema di NMT raggiungendo o avvicinandosi il più possibile al valore 0 (Monti 2019: 103).

Pérez-Ortiz et al. (2022: 148-149) illustrano i dettagli tecnici alla base di questo processo. Per avviare l'addestramento è necessario selezionare un insieme di pesi casuali o presi da una rete neurale che assolve a un compito simile. Questi pesi vengono modificati per minimizzare il più possibile il valore di una funzione di errore (nota anche come *loss function*, funzione di perdita), che quantifica la differenza tra output effettivi e output desiderati. Durante l'addestramento, si utilizzano gli "algoritmi di addestramento" (detti anche "algoritmi di apprendimento") che calcolano ripetutamente le piccole correzioni dei pesi fino a quando la funzione di errore risulta minima o sufficientemente piccola per tutti gli esempi del corpus di addestramento. L'approccio più comune è la "discesa del gradiente", che consiste nel calcolare in che modo la funzione di errore cambia al variare di ciascun peso secondo una quantità fissa ma molto piccola (il "gradiente della funzione di errore") e nel modificare i pesi nella direzione che riduce la funzione di errore. L'intensità delle variazioni dei pesi è controllata da un parametro chiamato "tasso di apprendimento", che si riduce progressivamente man mano che i pesi si avvicinano ai loro valori finali. Una volta stabilizzati i pesi, l'addestramento si conclude e la rete può essere impiegata per generare output a partire da nuovi input non usati precedentemente durante l'addestramento (Pérez-Ortiz et al. 2022: 148-149).

La fase di addestramento, che generalmente dura alcune settimane, si svolge in autonomia: i sistemi ricavano autonomamente le informazioni dai dati e imparano dagli errori. L'output generato sarà la migliore soluzione che emerge come risultato dei diversi tentativi, delle correzioni e delle elaborazioni dei dati disponibili (Monti 2019: 103).

Oltre a un corpus di addestramento, è necessario disporre anche di un piccolo corpus che non viene utilizzato per l'addestramento. Si tratta del *development corpus*, il cui scopo è di monitorare le prestazioni del sistema durante il *training* (Pérez-Ortiz et al. 2022: 161).

In fase di addestramento potrebbe verificarsi un problema: un *training* troppo approfondito sul corpus potrebbe compromettere la capacità di generalizzazione, poiché la rete neurale rischierebbe di “memorizzare troppo” le traduzioni di esempio. A causa di questo fenomeno (*overfitting*) il modello funziona perfettamente sul corpus di addestramento, mentre si adatta in modo non ottimale a un *test corpus*, composto da dati che la rete non ha mai visto prima e che è usato oppositamente per valutare le prestazioni del modello¹⁰. Ciò avviene perché il modello fatica a gestire le informazioni presenti nel *test corpus*, che differiscono da quelle presenti nel corpus di addestramento (Ying 2019: 1). In altre parole, il modello ottiene prestazioni elevate sui dati di *training*, ma scarse performance su dati nuovi risultando quindi non funzionale per lo scopo designato e poco performante.

Pertanto, il *development corpus* può essere utilizzato durante il processo di addestramento, prima del suo completamento e dopo un certo numero di fasi di *training* (epoche), per tradurre le sue frasi di partenza con la rete neurale e per confrontare l'output generato con le frasi di arrivo attese, ricorrendo a metriche di valutazione automatica approssimativa (Pérez-Ortiz et al. 2022: 161). La metrica più usata in questo caso è BLEU (Papineni et al. 2002). Questa calcola un punteggio che varia da 0 a 100%, dove 0 indica nessuna corrispondenza. Se il punteggio BLEU sul *development corpus* segnala un peggioramento delle prestazioni, l'addestramento può essere interrotto. In alternativa è possibile anche memorizzare l'attuale set di pesi e continuare l'addestramento per vedere se il punteggio migliora nuovamente (Pérez-Ortiz et al. 2022: 161). BLEU non è l'unica metrica di valutazione automatica che può essere impiegata per questo processo. Ulteriori metriche saranno approfondite nel capitolo 2.

1.3.2 Traduzione automatica adattiva

La traduzione automatica adattiva è una forma avanzata di TA, le cui prestazioni migliorano progressivamente grazie all'interazione continua con l'utente, sfruttando feedback come correzioni, glossari terminologici e valutazioni qualitative delle traduzioni. Questo approccio risulta particolarmente utile in contesti specifici, in quanto i sistemi di TA generici spesso non dispongono di dati sufficienti per trattare il linguaggio specialistico di interesse (Moslem et al. 2023). Pertanto, in questa nuova forma di TA, la macchina non svolge da sola il computo di traduzione, ma è

¹⁰ Il *test corpus* viene impiegato alla conclusione della fase di addestramento delle reti neurali con lo scopo di valutare come la rete si comporta con dati diversi da quelli impiegati per l'addestramento (Pustejovsky e Stubbs 2013).

possibile individuare un'interazione tra l'utente e la macchina in cui, come precisa Krüger (2019: 144), l'utente continua ad avere il controllo del processo traduttivo.

La TA adattiva si serve di diverse strategie di adattamento. Una prima è l'adattamento al dominio (*Domain Adaptation*), che permette al sistema di specializzarsi in un dominio specifico (ad esempio l'ambito medico o legale). Per attuare questa strategia si può eseguire un *fine-tuning*, ovvero un'ulteriore fase di addestramento di un modello neurale già pre-addestrato su un corpus specifico del dominio scelto (Saunders 2022: 12). Si impiega quindi una raccolta di testi rappresentativi del dominio in cui si intende specializzare il sistema permettendo allo stesso di apprenderne le peculiarità linguistiche¹¹. Sistemi come Google AutoML Translation offrono strumenti per addestrare modelli personalizzati su dati specifici forniti dall'utente.

Una seconda strategia è l'apprendimento in tempo reale (*Online Learning*), secondo il quale il sistema apprende istantaneamente da correzioni effettuate durante il *post-editing*. Questo paradigma di apprendimento automatico elabora i dati uno alla volta, aggiornando il modello a ogni nuovo esempio ricevuto (Peris e Casacuberta 2019: 2). Il processo può essere descritto in quattro fasi: presentazione al sistema di un campione, generazione della previsione, ricezione della risposta corretta e aggiornamento del modello sulla base di tale feedback (Murphy 2014). Il sistema utilizzerà la previsione corretta per adattare i suoi modelli. Un esempio di implementazione efficace di questa strategia è rappresentato da ModernMT, che integra l'apprendimento continuo durante la traduzione.

Un'ulteriore strategia di adattamento prevede l'uso di una memoria di traduzione, come avviene ad esempio nei CAT tools che permettono l'implementazione di sistemi di TA come Trados Studio, già citato in precedenza, o MateCat, un altro CAT tool reperibile gratuitamente online (Federico et al. 2014). Questa strategia si basa sul riutilizzo di traduzioni eseguite precedentemente grazie a un allineamento in tempo reale con le memorie di traduzione esistenti. Sono disponibili sistemi di TA commerciali che offrono questa funzionalità, uno di questi è ModernMT. Si tratta di un software di TA *open-source* che presenta una memoria dinamica interna che memorizza e si sincronizza con tutte le memorie di traduzione degli utenti (Bertoldi et al. 2018). ModernMT può essere integrato in MateCat: durante la traduzione, il software analizza il contesto, richiama gli esempi di traduzione più pertinenti dalla sua TM e adatta in tempo reale la rete neurale alla richiesta (Bertoldi et al. 2018). Un altro strumento è memoQ AGT (Adaptive Generative Translation) che impiega risorse linguistiche presenti come memorie di traduzione e glossari per generare traduzioni che si adattino immediatamente al dominio specifico dell'utente¹².

¹¹ Per una descrizione dettagliata della fase di *fine-tuning* eseguita nella presente ricerca si veda Capitolo 7.

¹² <https://www.memoq.com/product/memoq-agt/> (ultima consultazione: 20/12/2025)

La personalizzazione terminologica e i vincoli terminologici (*terminology constraints*) costituiscono un'ulteriore strategia di adattamento. Si tratta di imporre l'uso di termini predefiniti durante la TA attraverso glossari vincolanti o dizionari specialistici assicurando coerenza terminologica, ad esempio, per cliente o settore/dominio. Se le memorie di traduzione possono essere considerate come dati di addestramento pronti all'uso per l'adattamento, i database terminologici risultano più complessi da trattare e nel corso degli anni sono stati proposti diverse strategie per gestirli (Dinu et al. 2019: 3063). Il *constrained decoding* è una delle più rilevanti e viene illustrato in diversi studi, tra cui in Chatterjee et al. (2017) e Hokamp e Liu (2017). In breve, il *constrained decoding* prevede che, quando il sistema rileva che un termine nel testo di partenza corrisponde a una voce del glossario, esso viene forzato a impiegare la relativa traduzione indicata dal glossario al momento della decodifica (Dinu et al. 2019: 3063). Ciò significa che il *decoder* è sottoposto a vincoli che vengono applicati rigorosamente durante la fase di decodifica. Alcuni sistemi di TA, come DeepL, Amazon Translate e Microsoft Translator, permettono di integrare glossari personalizzati direttamente nel processo traduttivo.

Infine, si possono individuare le strategie di multi-dominio e meta-apprendimento. Il multi-dominio permette a un singolo sistema neurale di tradurre testi provenienti da più domini differenti (Saunders 2022: 5). Il meta-apprendimento, o *meta-learning*, invece, mira a creare modelli neurali in grado di apprendere in pochi passaggi di addestramento nuovi compiti sulla base di una quantità ristretta di nuovi dati (Finn et al. 2017). Tali strategie sono particolarmente utili per scenari in cui i dati di addestramento a disposizione sono limitati o i domini sono estremamente variabili.

1.4. *Augmented Translation*

Il concetto di *augmented translation* è stato introdotto da CSA Research, una società di analisi specializzata nel settore della localizzazione e della traduzione, in seguito alla pubblicazione di due articoli scritti da Lommel (2017) e DePalma (2017). DePalma e Lommel hanno delineato questo nuovo paradigma, ispirato al concetto di *augmented reality*, che rappresenta un approccio in continua evoluzione e che mira a potenziare il lavoro dei traduttori integrando in modo sinergico diverse tecnologie all'interno del processo traduttivo. In particolare, il CSA report descrive l'*augmented translation* come segue: “where artificial intelligence-driven technologies combine human and machine capabilities to enhance the effectiveness and efficiency of human translators” (cit. in. O'Brien 2024: 399). L'obiettivo pertanto è migliorare l'efficacia e l'efficienza del traduttore grazie a tecnologie basate sull'intelligenza artificiale.

Sebbene in letteratura l'*augmented translation* venga talvolta ridotta alla cooperazione tra TM e TA, con particolare enfasi sulla TA neurale (cfr. ad es. Lommel et al. 2017), autori come

Krüger (2019: 144) ne propongono una visione più estesa. Tale prospettiva abbraccia uno spettro tecnologico più vasto, considerando che la gamma di strumenti a disposizione del traduttore è in costante espansione ed evoluzione.

La rappresentazione dell'*augmented translation* proposta da CSA è una rappresentazione antropocentrica in cui lo specialista viene posto al centro ed è circondato da una serie di tecnologie (cfr. Fig. 5).

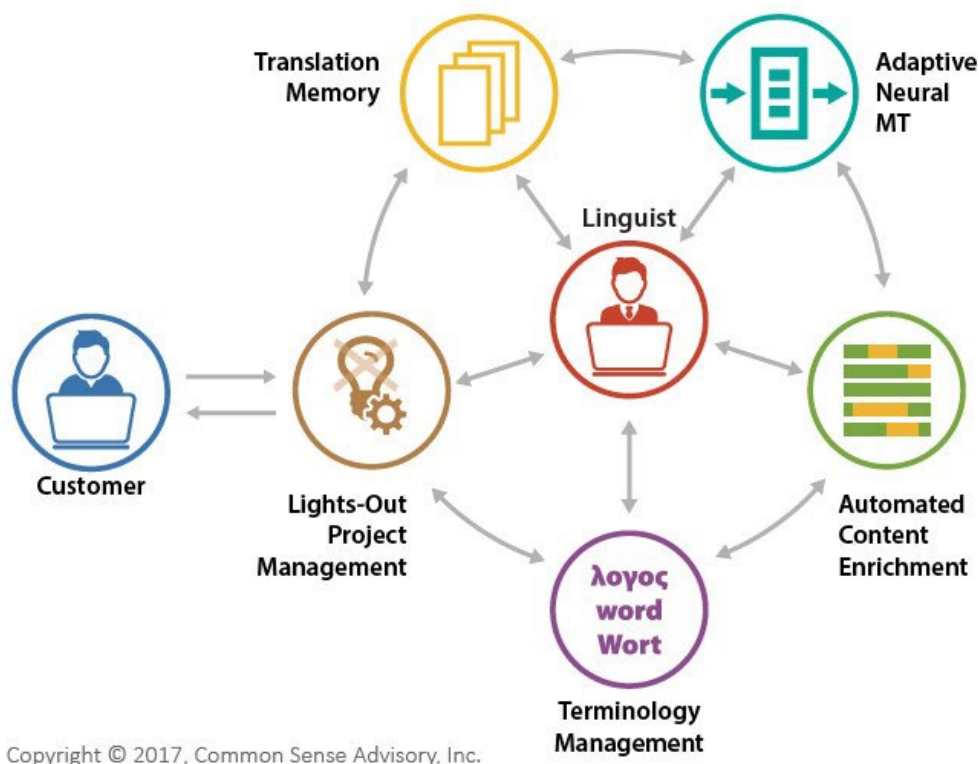


Figura 5: Augmented Translation puts Linguists in the center, DePalma (2017)

Nell'immagine è possibile individuare due concetti non ancora menzionati e descritti in questa sede: *Lights-Out Project Management* e *Automated Content Enrichment*. Il *Lights-Out Project Management* permette di automatizzare i processi di gestione dei progetti (come la fatturazione) che non richiedono l'intervento umano e che sottraggono del tempo lavorativo. L'*Automated Content Enrichment*, invece, è una tecnologia che consente di associare automaticamente i termini a risorse autorevoli, supportando l'utente nella disambiguazione e migliorando di conseguenza la TA (DePalma 2017).

Quest'ottica antropocentrica, delineata da CSA, illustra chiaramente il concetto di *Human-Centered Augmented Translation* introdotto da O'Brien (2024: 402), secondo cui il traduttore rimane al centro del processo, affiancato da tecnologie che ne potenziano e ampliano le capacità senza sottrargli il controllo decisionale. Il concetto di *augmented translation* sottolinea quindi il

ruolo attivo del traduttore, che beneficia del supporto tecnologico ma, come osserva Krüger (2019: 144), mantiene sempre il controllo completo sul processo traduttivo, pur collaborando strettamente con la macchina. Tenendo conto di quanto affermato, si può dedurre quindi che il concetto di *augmented translation* rappresenta e illustra l'interazione uomo-macchina nel lavoro di traduzione allo stato attuale nel modo più verosimile possibile.

1.5 Intelligenza artificiale e *Large Language Model*

A conclusione di questo capitolo e alla luce dei recenti sviluppi tecnologici, risulta fondamentale annoverare tra le tecnologie per la traduzione anche l'intelligenza artificiale (IA). Sebbene l'IA non sia oggetto del presente studio, il ruolo che svolge nella vita di tutti i giorni, per i professionisti e gli utenti comuni, è ormai innegabile. Pertanto, in questa sede se ne tratteranno i lineamenti essenziali, pur senza addentrarsi in una disamina approfondita che esulerebbe dagli obiettivi specifici di questa ricerca.

Il termine "intelligenza artificiale" è stato coniato da John McCarthy, considerato il padre dell'IA, nel 1956 in occasione del *Dartmouth Summer Research Project* (McCarthy et al. 2006). Attualmente esistono diverse definizioni di intelligenza artificiale che tentano di spiegare un termine il cui significato appare estremamente difficile da delimitare. Come dichiarato da Jiang et al. (2022), il focus dell'IA è la simulazione e l'ampliamento dell'intelligenza umana tramite teorie, metodi, tecnologie e applicazioni di ricerca che sono alla base dell'IA stessa. Tuttavia, occorre considerare che l'attuale campo di ricerca dell'IA costituisce un insieme eterogeneo di più campi di ricerca, ciascuno caratterizzato da obiettivi, metodologie e contesti applicativi specifici (Wang 2019: 28).

Attualmente, la maggior parte dei sistemi di IA afferisce alla categoria della cosiddetta *artificial narrow intelligence* (ANI) progettata per svolgere compiti specifici (Jiang et al. 2022). A essa si contrappone l'*artificial general intelligence* (AGI), che indica sistemi in grado di svolgere qualsiasi compito cognitivo tipico dell'essere umano, come comprendere, apprendere e risolvere problemi in maniera flessibile (Legg e Hutter 2007; Wang 2019).

In questo contesto, l'avvento dei *Large Language Models* (LLM) ha segnato una svolta fondamentale. Come suggerisce la denominazione, i LLM sono modelli linguistici di grandi dimensioni. Più precisamente, si tratta di algoritmi di *deep learning*. Basati su reti neurali artificiali e diverse architetture, questi modelli sono pre-addestrati su grandi quantità di dati testuali e sono in grado di operare, con diversi livelli di efficacia, in un pluralità di contesti applicativi: dalla comprensione e generazione del linguaggio naturale (Brown et al. 2020), al ragionamento e risoluzione di problemi matematici complessi (Rein et al. 2023; Wei et al. 2023), alla generazione di

codici (Chen et al. 2021; Liu et al. 2024), fino alla produzione di traduzioni in diverse lingue (Shi et al. 2022).

Gli evidenti e significativi progressi compiuti hanno contribuito a diffondere l'idea di un passaggio concretamente possibile dall'*artificial narrow intelligence* verso l'*artificial general intelligence*, sebbene, come evidenziano Kim et al. (2025) quest'ultima resti ancora una prospettiva futura.

Allo stato attuale dell'arte, i GPT (*Generative Pre-Trained Transformer – Transformer* generativi pre-addestrati) sono il tipo più noto di modello linguistico di grandi dimensioni. Il primo GPT è stato presentato dall'azienda americana di IA OpenAI (Yenduri et al. 2024). Come suggerisce il nome, i modelli GPT sono reti neurali artificiali basate sull'architettura *Transformer*; questi modelli rappresentano un importante progresso nel campo dell'IA in quanto consentono applicazioni di IA generativa, che viene spiegata come segue: “computational techniques that are capable of generating seemingly new, meaningful content such as text, images, or audio from training data” (Feuerriegel et al. 2024: 111). L'IA generativa può quindi generare vari contenuti come testi, codici, immagini, video e musica. Esempi di IA generativa sono Midjourney¹³, DALL-E¹⁴ e ChatGPT¹⁵. Questa architettura risulta particolarmente adatta allo svolgimento di una serie di compiti legati alla NPL: può ad esempio generare, completare e riassumere testi, analizzare e comprendere il significato di un testo, rispondere a domande. È in grado anche di generare codici in linguaggi di programmazione come Python e JavaScript (Yenduri et al. 2024). Tra i compiti legati al NPL che i GPT possono svolgere, compare anche la traduzione. È tuttavia necessario precisare che i GPT non costituiscono dei sistemi di TA in senso stretto, bensì sono dei generatori di testi. A differenza dei sistemi di TA propriamente detti, che sono addestrati specificamente su corpora paralleli e mirano a stabilire una corrispondenza sistematica tra testo di partenza e testo di arrivo, i GPT generano output sulla base di probabilità di co-occorrenza apprese durante l'addestramento su grandi quantità di dati eterogenei. Ne consegue che, pur potendo produrre traduzioni su richiesta dell'utente, il modello non garantisce un parallelismo frase per frase né una corrispondenza controllata tra le due lingue, operando piuttosto come un generatore di testi plausibili nella lingua di arrivo.

Concludendo, l'avvento delle tecnologie per la traduzione ha riconfigurato profondamente la professione del traduttore e le sue prassi operative. I CAT tools rappresentano ormai un ausilio imprescindibile per i traduttori, permettendo loro di essere più efficienti e performanti. I sistemi di TA e gli strumenti di IA a loro volta si stanno consolidando nel settore dei servizi linguistici,

¹³ <https://www.midjourney.com/home> (ultima consultazione: 20/12/2025)

¹⁴ <https://openai.com/index/dall-e-3/> (ultima consultazione: 20/12/2025)

¹⁵ <https://chatgpt.com/> (ultima consultazione: 20/12/2025)

ponendo tuttavia diversi interrogativi riguardanti la qualità, l'etica e il ruolo del traduttore umano. Nel prossimo capitolo si analizzerà in particolare la questione della qualità della traduzione generata da un sistema di TA e della sua valutazione.

Valutazione della qualità della TA

Il presente capitolo tratta il tema della valutazione della qualità della traduzione automatica. Inizialmente si forniscono definizioni fondamentali, successivamente vengono presentate le metriche di valutazione divise in due approcci: metriche automatiche e metriche umane. Il capitolo prosegue fornendo una panoramica completa delle diverse metriche e analizzandole nel dettaglio.

2.1 Concetti e definizioni

La valutazione della qualità della traduzione è da tempo al centro del dibattito negli studi traduttologici, coinvolgendo studiosi, educatori e professionisti in una riflessione articolata e continua (Han 2020). Con il progressivo sviluppo e l'ampia diffusione dei sistemi di TA, la questione della valutazione della qualità è diventata sempre più rilevante, come dimostrano numerosi studi (cfr. ad es. Bentivogli et al. 2018; González e Giménez 2014; Graham et al. 2017; Zhou et al. 2008). L'implementazione delle tecnologie per la traduzione, come i CAT tools e i sistemi di TA, esercita infatti un impatto diretto sulla qualità del prodotto finale (Drugan et al. 2023: 195), rendendo urgente una riflessione metodologica solida su come misurare la qualità in modo sistematico e affidabile. L'argomento riveste un ruolo fondamentale, in quanto valutare nel contesto della TA non significa solo confrontare le prestazioni di diversi sistemi, ma anche individuare aree di miglioramento per ottimizzare l'efficacia complessiva dei modelli impiegati (González e Giménez 2014).

Secondo White (2003: 213), la valutazione può essere intesa come un processo che consiste nel misurare una determinata caratteristica confrontandola con uno standard di riferimento relativo a quella stessa caratteristica. Lo standard viene indentificato con la traduzione “corretta”, tuttavia, il concetto di correttezza può risultare difficile da circoscrivere. Uno stesso testo può dare origine a molteplici traduzioni, poiché diversi traduttori possono proporre soluzioni differenti, la cui qualità può variare in misura significativa (White 2003: 213). Questi aspetti rendono quindi la valutazione della qualità delle traduzioni generate da sistemi di TA un processo piuttosto complesso.

Prima di procedere con l'analisi dei metodi di valutazione esistenti, è fondamentale definire il concetto di qualità della traduzione. Koby et al. (2014: 416) definiscono la qualità della traduzione umana come segue: “A quality translation demonstrates accuracy and fluency required for the audience and purpose and complies with all other specifications negotiated between the requester and provider, taking into account end-user needs.”. Tale definizione individua i criteri cardine della qualità nell'accuratezza, nella scorrevolezza, e nel rispetto delle eventuali richieste

specifiche del committente nonché delle esigenze dell'utente finale. A tal proposito, Chatzikoumi (2020) definisce la scorrevolezza come la correttezza grammaticale e la naturalezza nella lingua di arrivo, mentre l'accuratezza è intesa come l'equivalenza semantica e pragmatica tra il testo di partenza e il testo di arrivo.

In questa sede, la definizione della qualità della traduzione fornita da Koby et al. sopra citata viene impiegata in riferimento alla TA, regolata dagli stessi principi della traduzione umana.

2.2 Metriche di valutazione della qualità

È possibile individuare due approcci alla valutazione della qualità della TA: la valutazione automatica e la valutazione umana. Come affermato da Chatzikoumi (2020), le tecniche di valutazione automatica sono caratterizzate da un intervento umano limitato alla fase di progettazione dello strumento valutativo, mentre l'intero processo si svolge senza ulteriore coinvolgimento umano. Le tecniche di valutazione umana, invece, implicano un'interazione diretta e attiva da parte di valutatori nel momento stesso in cui la qualità viene valutata, indipendentemente da eventuali interventi avvenuti nelle fasi precedenti. È possibile individuare anche metriche di valutazione semi-automatica, di seguito inserite tra le tecniche di valutazione umana alla luce della definizione di quest'ultima già fornita da Koby et al. (2014).

Questo compito può essere svolto sia da valutatori professionisti che da valutatori non professionisti. Inoltre, il processo valutativo può essere eseguito in forma individuale, o da gruppi di valutatori. Nel caso di valutazione individuale, il giudizio viene formulato da un solo valutatore che analizza la qualità della traduzione seguendo criteri prestabiliti e fornendo un risultato univoco (Castilho et al. 2018: 23). Questo approccio garantisce un giudizio diretto e immediato, ma può essere influenzato da una forte soggettività. Se svolta in gruppo, la valutazione della qualità prevede, invece, il coinvolgimento di più valutatori, ciascuno dei quali fornisce una propria analisi e stima, indipendente da quella degli altri esaminatori. I risultati formulati dai singoli valutatori vengono successivamente aggregati e viene elaborata una media dei punteggi. Questo approccio ha l'obiettivo di ridurre l'impatto di eventuali *bias* individuali, favorendo una maggiore oggettività complessiva (Castilho et al. 2018: 23).

Spesso le valutazioni in forma di gruppo vedono un'ampia partecipazione di valutatori non professionisti. Questa scelta è dettata principalmente da considerazioni economiche: il reclutamento di gruppi di valutatori esperti comporta costi elevati (Castilho et al. 2018: 23). Di conseguenza, i valutatori non esperti vengono preferiti per ridurre i costi, accettando il compromesso consistente in un livello inferiore di competenza individuale. Tuttavia, secondo Castilho et al. (2018: 23), il coinvolgimento di valutatori non esperti è giustificato dall'idea che il numero di partecipanti possa,

almeno in parte, compensare il loro livello di competenza inferiore, garantendo comunque risultati utili e rappresentativi.

2.2.1 Tecniche di valutazione umana della TA

Di seguito vengono trattate le tecniche di valutazione della TA basate su un intervento umano. Chatzikoumi (2020) propone una classificazione che individua due principali categorie di valutazione umana: le metriche basate su giudizio direttamente espresso (*directly expressed judgment, DEJ-based*) e quelle che non si basano su tale giudizio (*non-DEJ-based*). Tale distinzione riflette le divergenze sostanziali tra i metodi impiegati dai valutatori per giudicare se una traduzione rispetti il contenuto del testo di partenza o/e se la lingua di arrivo sia scorrevole e grammaticalmente corretta. Di conseguenza, la classificazione si basa sulle specificità del tipo di compito richiesto al valutatore. In particolare, nei metodi *DEJ-based*, ai valutatori è chiesto di esprimere un'opinione sulla qualità della traduzione, ad esempio giudicandola come adeguata o meno, oppure valutandone la scorrevolezza e l'accuratezza. Le metriche *non-DEJ-based*, invece, si distinguono per un approccio più operativo basato su attività strutturate.

Come sottolinea Chatzikoumi (2020), dal punto di vista cognitivo, differiscono anche i processi coinvolti nei due metodi. Nelle metriche *DEJ-based*, i valutatori devono esprimere un giudizio sulla qualità del testo, un compito che richiede una valutazione globale e un'analisi di molteplici aspetti linguistici e semantici. Nei metodi *non-DEJ*, invece, l'elaborazione cognitiva è più focalizzata e orientata alla risoluzione di compiti definiti, come individuare errori specifici o rispondere a domande. Questa impostazione rende pertanto il processo di valutazione più sistematico. In sintesi, la scelta tra i due approcci dipende dall'obiettivo della valutazione: se l'obiettivo è ottenere una valutazione complessiva della qualità, le metriche *DEJ-based* sono più appropriate; se si intende invece ottenere un'analisi più sistematica e oggettiva degli errori, le metriche *non-DEJ-based* risultano più efficaci.

2.2.1.1 Metodi di valutazione *DEJ-based*

Le valutazioni *DEJ-based* si basano su un'interpretazione soggettiva e possono essere influenzate dalla tolleranza linguistica, dalla formazione e dall'esperienza individuale del valutatore, rendendo quindi il processo più influenzabile da variabili personali, non sempre controllabili. Sebbene esistano linee guida per ridurre la variabilità tra valutatori, la componente soggettiva resta centrale (Chatzikoumi 2020).

Chatzikoumi (2020) individua diversi metodi di valutazione *DEJ-based*. Quello più usato si basa sulle categorie di *adequacy* e di *fluency*, misurate con strumenti specifici. Un ulteriore

approccio di valutazione *DEJ-based* consiste nel *ranking*, ovvero la classificazione di diversi sistemi di TA mediante il loro confronto o la formulazione di un giudizio, generale o specifico, sulla traduzione. Rientra tra questi metodi la valutazione diretta (*direct assessment*) in cui si richiede ai valutatori di fornire un accertamento della qualità di un singolo output di TA rispetto a una singola traduzione di riferimento, utilizzando una scala analogica. Tali metodi vengono esaminati di seguito in maniera più dettagliata.

- *Adequacy e fluency*

Questo metodo prevede una valutazione sulla base dell'*accuracy* (accuratezza), anche chiamata *adequacy*, e della *fluency* (scorrevolezza), la cui valutazione viene effettuata confrontando il testo di partenza con il testo di arrivo o il testo di arrivo con una traduzione di riferimento¹⁶.

L'accuratezza riguarda il contenuto del testo di arrivo: ai valutatori viene richiesto di stabilire l'adeguatezza del contenuto del testo di arrivo rispetto alla traduzione di riferimento o del testo di partenza. Con il termine "scorrevolezza" si fa riferimento invece al linguaggio del testo di arrivo. Il compito dei valutatori prevede, pertanto, di stabilire se il linguaggio dell'output della TA risulti naturale, corretto grammaticalmente e comprensibile nella lingua di arrivo (Callison-Burch et al. 2007: 140).

Il Linguistics Data Consortium ha sviluppato delle scale di valutazione per l'*Annual NIST Machine Translation Evaluation Workshop* (LDC 2005). Per l'accertamento dell'adeguatezza, al valutatore si chiede quanto del significato della traduzione di riferimento sia espresso nella traduzione associando i seguenti valori: 5 = *All*; 4 = *Most*; 3 = *Much*; 2 = *Little*; 1 = *None*. Per la valutazione della *fluency*, invece, il valutatore indica quanto la traduzione sia scorrevole e, nel caso di una traduzione verso l'inglese, userà la seguente scala: 5 = *Flawless English*; 4 = *Good English*; 3 = *Non-native English*; 2 = *Disfluent English*; 1 = *Incomprehensible* (Callison-Burch et al. 2007: 140). La valutazione dell'*adequacy* richiede una competenza bilingue, nella lingua di partenza e nella lingua di arrivo, mentre la valutazione della *fluency* richiede solo competenze nella lingua di arrivo (Castilho et al. 2018: 18).

È possibile individuare diversi strumenti che supportano questo metodo valutativo. Uno di questi è Dynamic Quality Framework (DQF), una piattaforma sviluppata dalla Translation Automation User Society (TAUS) nel 2011 (Görög 2014). Un altro approccio è Human UCCA-Based MT Evaluation (HUME), proposto da Birch et al. (2016).

¹⁶ Con il termine "traduzione di riferimento" si intende la traduzione prodotta dal traduttore professionista, considerata corretta e autorevole e utilizzata come termine di paragone per valutare la qualità della TA presa in esame.

- Ranking

Il *ranking* è un metodo di valutazione che si basa su un confronto tra output generati da diversi sistemi di TA e provenienti dallo stesso testo di partenza. I valutatori ordinano le frasi oggetto della valutazione sulla base di criteri o concetti generali indicati, ad esempio la *fluency*. Questo approccio presenta il vantaggio di essere intuitivo e accessibile, sia dal punto di vista dell'interpretazione dei risultati, sia dal punto di vista della sua implementazione pratica (Castilho et al. 2018: 21).

Tuttavia, sebbene il *ranking* sia caratterizzato da un'apparente semplicità, esso presenta svantaggi rilevanti: “the relatively low annotator agreement rates, the immense amount of annotator time required, and the difficulty of scaling the sentence ranking task to many systems” (Bojar et al. 2016: 29). Tra le criticità principali figura il cosiddetto *inter-annotator agreement* (accordo tra annotatori), ovvero la coerenza tra i giudizi espressi da valutatori diversi. Nonostante l'impiego di linee guida condivise, si possono osservare discrepanze nella classificazione delle traduzioni, con conseguenti tassi di accordo relativamente bassi. Questo fenomeno solleva interrogativi sulla riproducibilità e affidabilità dei dati raccolti mediante questo tipo di valutazione. Inoltre, il compito di ordinare le traduzioni può risultare cognitivamente impegnativo e particolarmente oneroso in termini di tempo. Un'ulteriore limitazione è rappresentata dalla difficoltà di scalabilità del metodo: all'aumentare del numero di sistemi da confrontare, cresce il numero di confronti necessari, rendendo il processo poco sostenibile per esperimenti su larga scala.

Per facilitare lo svolgimento di questo procedimento valutativo, sono stati sviluppati strumenti specifici, come Appraise (Federmann 2010), un tool *open-source* con cui è possibile eseguire dei task di annotazione di valutazione della qualità della TA.

- Direct Assessment

Secondo Bentivogli et al. (2018: 62), il metodo di valutazione diretta (*direct assessment* - DA) prevede da parte del valutatore la formulazione di un giudizio su una traduzione che viene confrontata con il testo di partenza o una traduzione di riferimento sulla base di una scala che va da 0 a 100. Nel caso in cui il confronto avvenga tra la TA e la traduzione di riferimento, si parla di *reference-based* DA (*DA-ref*). Se il confronto invece viene effettuato con il testo di partenza, si parla di *source-based* DA (*DA-src*). La *reference-based* DA può introdurre *reference bias* dovuti alla dipendenza dalle traduzioni di riferimento. Per superare questo problema, si può fare ricorso alla *source-based* DA (Bentivogli et al. 2018: 62).

I risultati della *Conference on Machine Translation* (WMT) del 2017 (Bojar et al. 2017) hanno indicato la valutazione diretta come metrica di valutazione standard e hanno inoltre evidenziato che, sebbene questo metodo possa basarsi sia sull'*adequacy* che sulla *fluency*, all'atto

pratico esso si concentri principalmente sul trasferimento semantico (che corrisponde all'adeguatezza), mentre il trasferimento sintattico (o scorrevolezza) è risultato essere meno rilevante.

- *Quality-checking annotation tasks*

Questo metodo di valutazione si basa su un task che riguarda il controllo della qualità della traduzione e che si concretizza nell'assegnazione dei seguenti tag: *acceptable* (accettabile), *can easily be fixed* (può essere facilmente corretto) e *none of both* (nessuno dei due) (Federmann 2012). Come sostiene Chatzikoumi (2020), il valutatore può incontrare diversi livelli di difficoltà nella correzione di un segmento a seconda della presenza di fattori che influenzano la complessità di traduzione del testo di partenza, come il grado di ambiguità, la chiarezza del linguaggio e/o della terminologia impiegata.

2.2.1.2 Metodi di valutazione *non-DEJ-based*

Nei metodi di valutazione *non-DEJ-based* i valutatori non devono esprimere un giudizio diretto, ma viene richiesto loro di eseguire attività strutturate, come l'identificazione e la classificazione degli errori secondo tassonomie predefinite o il completamento di task specifici (*task-based*), o di svolgere attività che testano la comprensione di un testo tradotto automaticamente. In questi casi, l'unico elemento che implica un giudizio è, secondo Chatzikoumi (2020), l'atto stesso della correzione, che suggerisce che la traduzione non sia perfetta. A differenza delle metriche *DEJ-based*, questo approccio assicura maggiore oggettività, in quanto riduce il margine di interpretazione personale, incanalando l'attenzione del valutatore verso attività operative più strutturate e meno opinabili.

In questi metodi rientrano le metriche semiautomatiche, le valutazioni basate su task, classificazione, analisi e correzione degli errori presenti negli output della TA (Chatzikoumi 2020), che vengono spiegate di seguito.

- *Metriche semiautomatiche*

Le metriche semiautomatiche, note anche come valutazione *human-in-the-loop*, rappresentano una forma ibrida di valutazione della qualità della TA che combina l'efficienza delle metriche automatiche con l'intervento umano da parte di annotatori. Tra queste, rientrano HTER, HBLEU e HMETEOR, ovvero le versioni semiautomatiche delle corrispondenti metriche automatiche TER,

BLEU e METEOR¹⁷. Nel caso della metrica semiautomatica HTER, non si calcola il numero di operazioni che devono essere effettuate dal valutatore per trasformare l'output della TA nella traduzione di riferimento, come avviene per la metrica automatica TER; si crea, invece, una nuova traduzione di riferimento apportando il minor numero possibile di modifiche all'output della TA in modo tale che sia scorrevole come la traduzione di riferimento e che si mantenga un'equivalenza semantica (Snover et al. 2006: 224).

- *Valutazione task-based*

Questo procedimento valutativo riguarda la valutazione dell'efficienza dell'output della TA in riferimento allo svolgimento di un compito specifico (Chatzikoumi 2020). A differenza delle metriche tradizionali, che si basano sul confronto diretto tra l'output della TA e una o più traduzioni di riferimento, la valutazione *task-based* si concentra sull'effettiva funzionalità dell'output nel supportare un compito reale.

Un esempio è rappresentato dallo studio condotto da Ageeva et al. (2015) in cui è stata effettuata una valutazione della qualità della TA per assimilazione impiegando un task di *gap-filling*. Questo approccio si basa sul principio secondo cui la comprensione di un testo da parte di un lettore è correlata al numero di parole che il lettore è in grado di ripristinare correttamente nel testo. Il task richiede ai valutatori umani di reinserire nelle traduzioni di riferimento le parole precedentemente eliminate o oscurate (Ageeva et al. 2015: 138).

Tale metodo permette ai valutatori coinvolti di effettuare un accertamento indiretto del grado di corrispondenza concettuale tra il testo di partenza e l'output della TA, senza operare un'analisi diretta della qualità linguistica dell'output (Chatzikoumi 2020). Questa modalità di valutazione particolarmente utile per i casi in cui la comprensione del contenuto prevale sulla qualità stilistica della traduzione.

- *Classificazione e analisi degli errori*

Il metodo di valutazione umana più usato è la classificazione e l'analisi degli errori, in cui uno o più valutatori sono incaricati di esaminare l'output generato dalla TA confrontandolo con il testo di partenza. Una volta individuati gli errori, i valutatori li classificano e li analizzano sulla base di apposite linee guida o tassonomie. Allo stato attuale dell'arte, è possibile individuare molteplici linee guida di classificazione degli errori. Una delle più conosciute è *Multidimensional Quality*

¹⁷ Per maggiori informazioni sulle metriche automatiche TER, BLEU e METEOR, si veda il paragrafo 2.2.2.1.

Metrics (MQM)¹⁸. Nel paragrafo 3.4, viene presentato un approfondimento dedicato alle linee guida per la classificazione degli errori.

- *Post-editing*

Il *post-editing* è un metodo di valutazione umana ampiamente utilizzato. O'Brien (2011: 197) lo definisce come segue: "the correction of raw machine translated output by a human translator according to specific guidelines and quality criteria".

È possibile individuare i seguenti passaggi che caratterizzano l'esecuzione del *post-editing* confrontando il testo di partenza con l'output generato dalla TA: 1) individuazione degli errori di traduzione confrontando il testo di partenza con quello di arrivo; 2) individuazione degli errori linguistici nella lingua di arrivo; 3) correzione degli errori; 4) revisione del segmento modificato (Chatzikoumi 2020). Il *post-editing*, pertanto, si configura come un'estensione del metodo di classificazione e analisi degli errori precedentemente citato. La valutazione della qualità della traduzione mediante *post-editing* dipende dagli errori generati dalla macchina individuati dal *post-editor*. A questa tecnica valutativa è dedicato il capitolo 3, in cui saranno affrontati e approfonditi i rilevanti aspetti teorici e pratici.

In sintesi, la scelta tra i due approcci sopra descritti dipende dall'obiettivo della valutazione: se l'obiettivo è ottenere una valutazione complessiva della qualità, le metriche *DEJ-based* sono più appropriate; se si intende invece ottenere un'analisi più sistematica e oggettiva degli errori, le metriche *non-DEJ-based* risultano più efficaci.

Come sottolinea Chatzikoumi (2020), dal punto di vista cognitivo, differiscono anche i processi coinvolti nei due metodi. Nelle metriche *DEJ-based*, i valutatori devono esprimere un giudizio sulla qualità del testo, un compito che richiede una valutazione globale e un'analisi di molteplici aspetti linguistici e semantici. Nei metodi *non-DEJ*, invece, l'elaborazione cognitiva è più focalizzata e orientata alla risoluzione di compiti definiti, come individuare errori specifici o rispondere a domande. Questa impostazione rende pertanto il processo di valutazione più sistematico.

2.2.1.3 Vantaggi e svantaggi delle tecniche di valutazione umana della TA

Sanders et al. (2011) hanno individuato i motivi principali per i quali la valutazione umana della TA è più vantaggiosa rispetto alla valutazione automatica. In primo luogo, poiché le traduzioni sono destinate all'uso umano e hanno come destinatario finale l'essere umano, la valutazione umana risulta la modalità più adeguata. In secondo luogo, le metriche automatiche non possiedono la

¹⁸ Per maggiori informazioni si veda <https://themqm.org/> (ultima consultazione: 20/12/2025)

comprensione del mondo che caratterizza e distingue gli esseri umani e che permette all'uomo di valutare adeguatamente la rilevanza degli errori di traduzione. Infine, il giudizio umano è considerato il riferimento per la qualità della traduzione poiché si ritiene non ci sia un sostituto adeguato in questo ambito. Le motivazioni suddette mettono in luce i vantaggi che il valutatore umano possiede in quanto essere umano: egli è una persona pensante con competenze specialistiche adeguate a svolgere il compito a lui assegnato.

Per quanto riguarda gli svantaggi della valutazione umana, si può individuare l'aspetto della soggettività, già precedentemente citato. In queste tecniche, infatti, il giudizio può differire da quello di altri valutatori. Ulteriori svantaggi riconosciuti da Dorr et al. (2011) sono i seguenti: il costo elevato richiesto dal valutatore umano che deve avere competenze in entrambe le lingue prese in esame durante la valutazione e che quindi lo rende più costoso; la mancanza di ripetibilità della valutazione; il carattere dispendioso in termini di tempo, in quanto il processo valutativo, essendo svolto da un essere umano, richiede necessariamente più tempo rispetto alla valutazione automatica; lo scarso accordo tra gli annotatori. Quest'ultimo problema, già precedentemente accennato, può essere gestito e affrontato imponendo criteri rigorosi e formando adeguatamente i valutatori (Chatzikoumi 2020).

Nonostante i notevoli progressi compiuti nella ricerca sulla valutazione umana della TA, permangono importanti interrogativi metodologici che incidono sulla qualità e sull'affidabilità dei risultati. Come sottolineato da Lommel, Popovic, et al. (2014), restano aperte le questioni riguardanti quante traduzioni di riferimento, classificazioni o quanti interventi di *post-editing* siano necessari per ogni frase al fine di formulare una valutazione. La mancanza di consenso su questi parametri metodologici mette in discussione la possibilità di ottenere risultati che siano al tempo stesso affidabili e replicabili. Un'ulteriore limitazione è segnalata da Läubli et al. (2018) e riguarda in particolare la valutazione umana svolta a livello di singole frasi. Questo approccio frammentato può essere ricondotto in parte alla natura operativa dei compiti valutativi, che spesso isolano le frasi per facilitare il confronto diretto con le traduzioni di riferimento. Gli autori evidenziano come questo tipo di valutazione fornisca giudizi incompleti in quanto non si tiene conto della coesione e della coerenza globale del testo. Pertanto, per evitare che vengano trascurati i difetti rilevati, sarebbe meglio impiegare un approccio olistico, prendendo in esame interi documenti o sequenze di frasi in contesto, al fine di cogliere eventuali incoerenze, ripetizioni inappropriate o passaggi poco fluidi che non emergano a livello della singola frase.

Nella scelta degli approcci di valutazione umana della TA, può essere utile tenere conto di tre fattori fondamentali individuati da Läubli et al. (2020) che conseguentemente possono influenzare

l'accertamento stesso: la scelta dei valutatori, l'uso del contesto linguistico e la costruzione delle traduzioni di riferimento. I tre aspetti vengono di seguito esaminati singolarmente.

- Scelta dei valutatori: traduttori esperti potrebbero fornire valutazioni più accurate rispetto a valutatori non-esperti e, quindi, le valutazioni di esperti potrebbero mettere in luce maggiori differenze tra gli output della TA e le traduzioni umane (Läubli et al. 2020).
- Contesto linguistico: la qualità della TA è stata valutata quasi sempre a livello frasale, poiché la maggior parte dei sistemi di TA opera a questo livello. Tuttavia, una valutazione limitata a singole frasi non consente di identificare errori legati alla coerenza e alla coesione testuale. Valutare interi documenti, invece, permette di cogliere più efficacemente incoerenze, ripetizioni inappropriate o passaggi poco scorrevoli, sfruttando il contesto interfrasale (Läubli et al. 2020).
- Traduzioni di riferimento: le traduzioni umane di riferimento con cui vengono confrontati gli output della TA svolgono un ruolo importante. Se si usano testi tradotti, è importante considerare che questi sono solitamente più semplici dei corrispondenti testi originali (Laviosa-Braithwaite 1998) e quindi dovrebbero essere più facili da tradurre per i sistemi di TA. Inoltre, diverse traduzioni umane dello stesso testo di partenza possono presentare notevoli differenze di qualità e il confronto con un sistema di TA ha senso solo se le traduzioni umane di riferimento sono di alta qualità. Considerando l'influenza che la qualità delle traduzioni umane può avere sui risultati della valutazione, è quindi necessario effettuare accurati controlli di qualità (Läubli et al. 2020).

Gli aspetti sopra delineati evidenziano come la valutazione umana della traduzione automatica sia un processo complesso e influenzato da molteplici fattori. A partire da queste considerazioni, Läubli et al. (2020) hanno formulato alcune raccomandazioni volte a rendere più solido ed efficace il procedimento della valutazione umana:

1. scegliere come valutatori dei traduttori professionisti;
2. valutare i documenti interi e non le singole frasi: la valutazione in ordine randomico di queste ultime impedisce di individuare errori di coerenza e coesione testuale;
3. valutare la scorrevolezza oltre all'accuratezza;
4. non modificare in modo eccessivo le traduzioni per aumentare la scorrevolezza: modifiche molto invasive possono rendere le traduzioni più scorrevoli ma meno accurate;
5. usare testi di partenza originali.

Gli aspetti rilevanti citati nelle raccomandazioni sono la scelta dei valutatori, il contesto linguistico, i criteri di valutazione, le modalità d'intervento dei valutatori e le traduzioni di riferimento.

Pertanto, queste indicazioni tengono conto degli aspetti critici della valutazione umana della TA, così come dei fattori che possono influenzarla.

2.2.2 Tecniche di valutazione automatica della TA

Chatzikoumi (2020: 140) definisce come segue le tecniche di valutazione automatica: “the systems which score MT [Machine Translation] outputs without any human involvement”. Ciò significa che le metriche di valutazione automatica consentono una valutazione dell’output senza la necessità di un intervento umano diretto nella fase valutativa. Tuttavia, il contributo umano resta necessario nella fase di impostazione del compito, ad esempio durante la raccolta dei dati, le annotazioni o la produzione di traduzioni di riferimento (Chatzikoumi 2020: 140).

Secondo Koehn (2009: 220-221), le metriche di valutazione automatica devono soddisfare quattro requisiti fondamentali: essere economiche, ovvero meno costose e dispendiose in termini di tempo rispetto alla valutazione umana; essere coerenti, producendo risultati affidabili e ripetibili; fornire valori significativi a prima vista, cioè facilmente interpretabili; e classificare in modo adeguato l’output della TA secondo una scala valutativa il più possibile accurata.

Considerando lo stato attuale dell’arte, si possono individuare tre principali tipologie di valutazione automatica: (i) metriche che si basano su traduzioni di riferimento; (ii) metriche di *confidence* o *quality estimation* (QE); (iii) valutazione diagnostica basata su *check-point* (Chatzikoumi 2020).

2.2.2.1 Metriche basate su traduzioni di riferimento

Le metriche basate su traduzioni di riferimento assegnano un punteggio all’output della TA che si basa sul grado di somiglianza con le traduzioni di riferimento, anche chiamate *gold translation*. Rientrano in questa categoria le metriche che si basano su *edit distance*, e *precision* e *recall*.

- *Edit distance*

Una sottocategoria di metriche che può essere ricondotta alle metriche basate su traduzioni di riferimento impiega l’*edit distance*, chiamata anche distanza di Levenshtein (Levenshtein 1966). Questa misura quantifica il numero minimo di operazioni necessarie per trasformare una stringa *a* in una stringa *b* (Navarro 2001). Nel contesto della TA, l’*edit distance* viene utilizzata per calcolare quante e quali modifiche devono essere apportate all’output della TA affinché corrisponda alla traduzione di riferimento. Levenshtein (1966) individua come operazioni di *edit distance* l’inserimento, l’eliminazione e la sostituzione di un carattere con un altro.

Una metrica di *edit distance* è la *Word Error Rate* (WER) (Nießen et al. 2000), comprese le sue variazioni ed estensioni. Essa si basa sull'insieme di operazioni di eliminazione, sostituzioni e inserimento effettuate tra traduzione proposta “t” e la traduzione di riferimento “r” (Nießen et al. 2000). Tra le sue varianti ed estensioni più famose è possibile individuare: *WERg* (Blatz et al. 2004), *Translation Edit Rate* (TER) (Snover et al. 2006) e *Inversion WER* (invWER) (Leusch et al. 2003). Il funzionamento di queste varianti ed estensioni è analogo a quello di WER.

La metrica *Multiple Reference WER* (MWER) (Nießen et al. 2000) estende la metrica originale utilizzando più traduzioni di riferimento. La metrica *Sentence Error Rate* (SER), invece, indica la percentuale di frasi per le quali la traduzione generata dalla macchina non corrisponde esattamente a nessuna delle traduzioni di riferimento (Tomás et al. 2003). Infine, HyTER è una versione automatizzata di HTER (Snover et al. 2006), una metrica semiautomatica citata precedentemente nel paragrafo 2.2.1.2.

- *Precision e recall*

Precision (precisione) e *recall* (richiamo) rappresentano tecniche ampiamente utilizzate per la valutazione della TA. Con il termine “precisione” si intende il rapporto tra le parole accettabili nell'output della TA (cioè le parole presenti in almeno una delle traduzioni di riferimento) e il numero totale di parole ricorrenti nello stesso output della TA. Il richiamo, invece, è il rapporto tra le parole accettabili nell'output della TA e il numero di parole della traduzione di riferimento (Chatzikoumi 2020). Pertanto, in una frase composta da 12 parole di cui 6 sono accettabili, la precisione è di 6/12. Per la stessa frase, invece, se la traduzione di riferimento ha 16 parole, il richiamo è di 6/16.

Rientra nelle metriche basate sulla precisione la *Bilingual Evaluation Understudy* (BLEU), introdotta da Papineni et al. (2002). Essa si basa sul seguente principio: “The closer a machine translation is to a professional human translation, the better it is.” (Papineni et al. 2002: 311). Pertanto, il concetto alla base di questa metrica è la vicinanza dell'output generato dalla TA alle traduzioni di riferimento. Per valutare l'output si usano i criteri di adeguatezza e scorrevolezza: in seguito a un confronto con le traduzioni di riferimento, viene effettuata una valutazione assegnando un valore su una scala da 0 a 1, dove 0 rappresenta il punteggio minimo. Pertanto, i due elementi indispensabili per la metrica BLEU sono il concetto di “translation closeness” (Papineni et al. 2002: 311) e un corpus di traduzioni umane di riferimento di buona qualità.

BLEU è ispirata alla metrica WER precedentemente citata, distinguendosi però da quest'ultima per l'uso di diverse traduzioni di riferimento e per la considerazione delle differenze legittime nella scelta e nell'ordine delle parole. BLEU, in particolare, si basa sull'impiego di una

media ponderata delle corrispondenze tra frasi di lunghezza variabile e le traduzioni di riferimento (Papineni et al. 2002). Su BLEU si basa NIST, il cui nome deriva dal US National Institute for Standards and Technology (Doddington 2002). Si tratta di un adattamento di BLEU che differisce per un aspetto fondamentale: invece di considerare il punteggio di precisione, si valuta maggiormente il guadagno informativo. In particolare, le corrispondenze di parole più rare ricevono un peso maggiore, poiché la loro corretta traduzione contribuisce in misura maggiore all'informazione complessiva, mentre le parole più comuni vengono penalizzate (Zhang et al. 2004).

Numerose altre metriche combinano la precisione e il richiamo, tra le quali: la *F-measure metric* (Melamed et al. 2003), le implementazioni di *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE) (Lin e Och 2004), che sono degli adattamenti di BLEU, la famiglia di metriche *Broad Learning and Adaptation for Numeric Criteria* (BLANC) (Lita et al. 2005) e *Length Penalty, Precision, n-gram Position difference Penalty and Recall* (LEPOR) (Han et al. 2012). Una delle metriche maggiormente note è METEOR, che valuta la traduzione calcolando un punteggio sulla base di corrispondenze esplicite parola-parola tra la traduzione generata dal sistema di TA e la traduzione di riferimento. Se sono disponibili più traduzioni di riferimento, la traduzione data viene valutata rispetto a ogni riferimento in modo indipendente e viene riportato il punteggio migliore (Banerjee e Lavie 2005: 66).

2.2.2.2 Confidence o quality estimation

Quality estimation (QE), anche definita come *confidence estimation*, rappresenta un approccio alternativo alla valutazione automatica della qualità delle traduzioni, in quanto non richiede il confronto con traduzioni di riferimento (Bojar et al. 2016: 31). A differenza delle metriche tradizionali, la QE permette di prevedere per un dato input la qualità dell'output di un sistema senza alcuna informazione sull'output atteso (Specia et al. 2010: 40). Questo la rende particolarmente utile per i contesti traduttivi reali, in cui le traduzioni di riferimento non sono disponibili.

Le prime metriche di QE applicate alla TA lavoravano inizialmente a livello di parole (Gandrabur e Foster 2003; Ueffing e Ney 2005). Successivamente sono state elaborate metriche orientate a livello di frase (Quirk 2004), un'evoluzione che riflette l'esigenza crescente di offrire ai professionisti strumenti più flessibili e più precisi.

Le moderne metriche di QE sono strumenti sviluppati e addestrati utilizzando algoritmi di apprendimento automatico supervisionato, basati su esempi di traduzioni previamente etichettate in base alla loro qualità. Le previsioni effettuate sulla base delle metriche di QE possono essere fornite a diversi livelli di granularità: parola, frase, paragrafo o documento intero. A seconda del livello di

granularità desiderato, variano le caratteristiche considerate, le tipologie di etichette utilizzate e gli algoritmi impiegati per la realizzazione dei modelli predittivi (Bojar et al. 2016: 31).

Un esempio significativo di implementazione della QE è rappresentato da QuEst, una piattaforma *open-source* sviluppata da Specia et al. (2013), articolata in due moduli principali indipendenti: uno di estrazione delle caratteristiche e uno di apprendimento automatico. Il primo modulo fornisce una serie di estrattori delle seguenti caratteristiche: *fluency*, *adequacy* e *confidence*. Per misurare queste caratteristiche, QuEst opera nel seguente modo: dai segmenti tradotti può estrarre le caratteristiche utili a misurare la scorrevolezza (*fluency*) delle traduzioni; dal confronto tra i segmenti di partenza e quelli di arrivo, può estrarre caratteristiche di adeguatezza (*adequacy*) volte a verificare se struttura e significato del testo di partenza siano stati mantenuti nella traduzione; inoltre, se disponibili, i dati forniti dal sistema di TA impiegato per produrre le traduzioni possono indicare il grado di affidabilità (*confidence*) del sistema stesso. Il modulo di apprendimento automatico, invece, utilizza le caratteristiche estratte dal testo di partenza e da quello di arrivo per creare modelli di QE mediante l'uso di algoritmi di regressione e classificazione (Specia et al. 2013: 80-82).

2.2.2.3 Valutazione basata sui *checkpoints*

Zhou et al. (2008) propongono un altro metodo di valutazione che si basa sui cosiddetti *checkpoints*, intesi dagli autori come unità linguistiche, quali parole ambigue, collocazioni verbo-oggetto o sintagmi preposizionali, definite precedentemente in una tassonomia utile alla valutazione. A differenza delle metriche tradizionali che forniscono un punteggio globale, la valutazione tramite *checkpoints* consente di isolare e diagnosticare la performance del sistema di TA rispetto a elementi linguistici critici.

Zhou et al. (2008: 1122) illustrano i passaggi che compongono questa tecnica di valutazione: in primo luogo, si crea un corpus di frasi parallele; in seguito, viene effettuata un'analisi grammaticale delle frasi di partenza e d'arrivo; successivamente si esegue l'allineamento per ogni frase; per ogni categoria di *checkpoint*, vengono estratte le unità linguistiche dalle coppie di frasi; infine, si individuano i riferimenti di tali unità linguistiche nelle frasi di partenza. La valutazione viene effettuata selezionando frasi di prova in base alle categorie di *checkpoint* da analizzare e per ogni *checkpoint* si calcola il numero di parole che coincidono con le traduzioni di riferimento nella traduzione proposta dal sistema di TA. Si calcola infine un punteggio per ogni categoria, che viene successivamente sommato per ottenere un risultato complessivo del sistema.

Questo tipo di metrica ha il vantaggio di permettere agli sviluppatori di valutare non solo l'efficacia complessiva del sistema, ma anche di individuare i punti di forza e di debolezza,

fungendo così da utile strumento diagnostico. Tuttavia, la sua efficacia dipende dalla qualità e dalla completezza della tassonomia linguistica utilizzata. In sintesi, la valutazione basata su *checkpoint* pone l'attenzione sulla competenza linguistica del sistema in aree linguisticamente sensibili.

2.2.2.4 Vantaggi e svantaggi delle tecniche di valutazione automatica della TA

Le metriche di valutazione automatica della qualità delle traduzioni offrono molteplici vantaggi, ma presentano anche alcune limitazioni che ne condizionano l'affidabilità e l'applicabilità. Esse si distinguono da quelle umane per la ridotta necessità di intervento diretto da parte dell'uomo, rappresentando un primo vantaggio significativo. Questo aspetto le rende meno costose, poiché elimina il problema di dover ricorrere a valutatori umani con competenze bilingui, spesso onerosi. Inoltre, al concetto dell'automaticità si associa anche a un maggiore oggettività (Castilho et al. 2018: 26). A ciò si aggiunge il vantaggio della coerenza: a parità di input e parametri, lo stesso testo ottiene sempre lo stesso punteggio, garantendo una riproducibilità difficilmente raggiungibile nelle valutazioni umane, spesso influenzate da variabilità individuale (Koehn 2009: 221).

Le metriche automatiche presentano un ulteriore vantaggio fondamentale: la rapidità con la quale si ottiene un risultato. Per questo motivo, vengono spesso impiegate durante lo sviluppo di sistemi di TA per testare le modifiche apportate e raccogliere indicazioni utili per orientarne in maniera ottimale lo sviluppo e potenziarne le prestazioni (Banerjee e Lavie 2005: 65).

Tuttavia, l'apparente oggettività delle metriche automatiche va considerata con cautela. Anche se l'elaborazione del punteggio è affidata a un sistema, le metriche automatiche non possono prescindere del tutto dai traduttori umani: questi ultimi, infatti, assumono un ruolo fondamentale nel processo di valutazione automatica di metriche basate su traduzioni di riferimento, in quanto responsabili dell'elaborazione delle stesse. Questo aspetto introduce anche in questo caso il fattore soggettività, anche se in misura più limitata (Castilho et al. 2018: 26). Va inoltre considerato che un testo o una frase di partenza possono avere diverse traduzioni corrette; pertanto, il confronto dell'output con una sola versione rischia di introdurre un margine di errore, penalizzando potenzialmente soluzioni altrettanto appropriate. Pertanto, come affermano Castilho et al. (2018: 26), seppure le metriche automatiche elaborino punteggi che assumono la forma di un giudizio affidabile e oggettivo, il loro funzionamento si basa su una serie di presupposti che sollevano dubbi sulla loro effettiva oggettività e validità.

Un ulteriore limite è rappresentato dalla scarsa capacità delle metriche automatiche tradizionali di valutare aspetti più complessi, quali l'equivalenza semantica o la struttura sintattica dell'output (Castilho et al. 2018: 26). Tali strumenti, infatti, non possiedono una reale capacità di analisi e comprensione linguistica, operando frequentemente su principi statistici. Per quanto

riguarda i modelli di *quality estimation*, Sun et al. (2020) osservano che è più probabile che essi tentino di indovinare la qualità della traduzione piuttosto che valutarla accuratamente, in quanto ignorano le relazioni semantiche tra i segmenti tradotti automaticamente e quelli originali e, pertanto, non sono in grado di formulare giudizi corretti sull'adeguatezza. Per ovviare a tali criticità, la ricerca recente si è orientata verso lo sviluppo di metriche più avanzate, che integrano elementi di analisi linguistica più sofisticata e tentano di colmare il divario tra valutazione automatica e umana (Castilho et al. 2018: 26).

In sintesi, le metriche automatiche si configurano come strumenti utili per lo sviluppo e il monitoraggio continuo dei sistemi di TA grazie alla loro velocità, economicità e coerenza. Tuttavia, non possono sostituire del tutto il giudizio umano. La creazione delle frasi di riferimento richiede l'intervento umano, che comunque introduce una componente soggettiva anche nelle valutazioni automatiche, ridimensionando l'idea che "automatico" sia sinonimo di "oggettivo". Di conseguenza, l'oggettività delle metriche automatiche è solo parziale, e il concetto stesso di "qualità" nella traduzione resta, almeno in parte, ancorato a interpretazioni e scelte soggettive.

Concludendo, le valutazioni umane, pur essendo più accurate sotto il profilo linguistico, risultano costose, soggettive e difficilmente scalabili. Le metriche automatiche, invece, si distinguono per rapidità, economicità e coerenza nei risultati, ma mostrano limiti nella capacità di cogliere aspetti semantici e stilistici complessi. La riflessione sui diversi approcci di valutazione della qualità della traduzione automatica presentata in questo capitolo risulta particolarmente rilevante in vista del capitolo successivo, in cui sarà oggetto di approfondimento il processo di *post-editing*.

Post-editing

Il presente capitolo si propone di illustrare in modo approfondito uno degli approcci di valutazione della qualità della TA: il *post-editing*. Dopo aver fornito una definizione generale, il capitolo analizza i diversi livelli di intervento possibili, distinguendo tra *light* e *full post-editing* e illustrando linee guida aggiornate per ciascuna modalità. Successivamente, si analizzano le differenze tra *post-editing* monolingue e bilingue, per poi soffermarsi sulla norma ISO 18587, uno standard internazionale. Segue una sezione dedicata alle linee guida di classificazione degli errori. Infine, si presentano approfondimenti riguardanti la retribuzione del *post-editor*, le competenze richieste per svolgere questa attività in modo efficace e, infine, le principali sfide che ne caratterizzano l'attuazione concreta.

3.1 Concetti e definizioni

Con il termine “post-editing” s'intende la revisione da parte di un traduttore umano di un testo tradotto automaticamente, ovvero generato da un sistema di TA. Veale e Way (1997) definiscono il *post-editing* come “term used for the correction of machine translation output by human linguists/editors”, sottolineando il ruolo fondamentale dell'intervento umano di correzione. Tuttavia, questa definizione si concentra quasi esclusivamente sull'atto della “correzione”, trascurando la complessità del processo cognitivo che esso comporta.

A tal riguardo, Loffler-Laurian (1985) propone una prospettiva più ampia, descrivendolo come un processo di modifica piuttosto che di revisione. Tale distinzione è significativa, in quanto non è solo terminologica, bensì anche concettuale. In primo luogo, come evidenziato già dalla definizione di Veale e Way (1997), il *post-editing* riguarda la correzione di un testo generato da un sistema computerizzato; la revisione, diversamente, comporta la correzione di una traduzione prodotta da un traduttore umano. Di conseguenza, i tipi di errori da correggere in questi due processi sono differenti: ad esempio, i sistemi di TA tendono a confondere le frasi affermative e quelle negative, errore inusuale tra i traduttori umani. Inoltre, come sostengono Macken et al. (2022), il *post-editing* e la revisione possono essere considerati due processi cognitivi differenti, in quanto il primo si configura come un processo bilingue, mentre la seconda è più frequentemente un processo monolingue incentrato sulla lingua di arrivo. Ulteriori fattori da considerare sono il tempo a disposizione per l'esecuzione e la qualità finale della traduzione: il *post-editing* viene commissionato ed effettuato generalmente perché si dispone di una quantità di tempo limitata. In

questa sede si fa riferimento alle definizioni proposte da Veale e Way (1997) e Loffler-Laurian (1985) sopra riportate che consentono di inquadrare concettualmente il termine.

Per spiegare il ruolo della qualità nel processo di *post-editing*, accennata come fattore di differenziazione tra il *post-editing* e la revisione, si è deciso di avvalersi della definizione fornita dal Report di TAUS¹⁹ del 2010, in cui il *post-editing* viene definito come segue: “the process of improving a machine-generated translation with a minimum of manual labor” (TAUS 2010, cit. in Screen 2017: 317). L’idea alla base di questa definizione è di intervenire il meno possibile sull’output della TA, e quindi di sfruttarlo al massimo delle possibilità. Alla luce di tali premesse, si può dedurre che la qualità della traduzione post-editata si differenzierà dalla qualità della traduzione prodotta ex novo da un traduttore umano, in quanto la traduzione post-editata manterrà un’impronta stilistica riconducibile all’origine automatica della traduzione. In quest’ottica, il *post-editing* mira a valorizzare il più possibile l’output della TA, accettando una qualità che si discosta da quella di una traduzione umana integrale.

Un’ulteriore definizione utile a introdurre un concetto chiave che guida il *post-editor* nello svolgimento del compito a lui assegnato è presentata da TAUS/CNGL²⁰ (2011, cit. in Rico Pérez e Torrejón 2012: 168) secondo cui il *post-editing* è la “correction of machine-generated translation output to ensure it meets a level of quality negotiated in advance between client and post-editor”. Quest’ultima definizione introduce una dimensione fondamentale assente nelle precedenti: il concetto di qualità negoziata. In questo caso, il *post-editing* non è descritto solo come un’attività tecnica o linguistica, ma come un processo vincolato da obiettivi predefiniti, stabiliti a monte in base alle esigenze del committente.

In questa sede, in riferimento alla delineazione teorica del concetto di *post-editing*, si fa riferimento alle due definizioni fornite da TAUS (2010) e TAUS/CNGL (2011) ponendole in una relazione in cui si condizionano reciprocamente. Coerentemente con l’approccio delineato da TAUS (2010), si ritiene che l’intervento del *post-editor* sull’output della TA debba essere il più contenuto possibile. Tuttavia, e in accordo con quanto espresso nella definizione di TAUS/CNGL (2011), si considera necessario che, nell’esecuzione del *post-editing*, il professionista incaricato prenda in considerazione le specifiche esigenze del cliente e che apporti le modifiche necessarie affinché il testo raggiunga il livello qualitativo tale da soddisfare quanto richiesto dal committente. Di conseguenza, il grado di intervento sul testo tradotto automaticamente dipenderà dalla qualità attesa: in alcuni casi sarà minimo, in altri più consistente. In questa prospettiva, l’idea di

¹⁹ TAUS (*Translation Automation User Society*) è un’organizzazione internazionale attiva nel settore della valutazione della qualità e del *post-editing* automatico. Per maggiori informazioni si veda: <https://www.taus.net/> (ultima consultazione: 20/12/2025).

²⁰ TAUS/CNGL indica la collaborazione tra TAUS (*Translation Automation User Society*) e il CNGL (*Centre for Next Generation Localisation*) che ha portato, nel 2011, alla produzione di documenti di riferimento per il *post-editing*.

“intervenire il meno possibile” risulta quindi relativa e soggetta a differenti interpretazioni in base al contesto e agli obiettivi prefissati.

La definizione di TAUS/CNGL (2011), inoltre, introduce la figura del *post-editor*, ovvero il professionista incaricato di intervenire sull’output grezzo prodotto dal sistema. Come descritto da Allen (2003: 297), “the task of the post-editor is to edit, modify and/or correct pre-translated text that has been processed by an MT system from a source language into (a) target language(s)”. In tale definizione, il compito del *post-editor* viene descritto in termini piuttosto operativi, quasi meccanici. Tuttavia, questa visione rischia di ridurre il *post-editor* a un mero “correttore” subordinato all’output tecnologico, senza riconoscere la complessità cognitiva e decisionale del suo lavoro²¹.

In conclusione, le definizioni analizzate nel presente paragrafo evidenziano diverse sfumature del concetto di *post-editing*, ma concordano nel riconoscere la centralità del lavoro umano nel garantire un livello di qualità adeguato del testo tradotto automaticamente.

3.2 I livelli di *post-editing*

Il *post-editing* si articola generalmente in due livelli distinti di intervento:

1. *light post-editing*
2. *full post-editing*

Il *light post-editing* può essere descritto come un intervento poco invasivo di modifica del testo tradotto dalla macchina, in cui l’attività del *post-editor* è finalizzata alla correzione degli errori essenziali. Si tratta di interventi mirati a correggere quelle imprecisioni che alterano il significato del testo e che potrebbero compromettere la corretta comprensione e la trasmissione delle informazioni al destinatario (Monti 2019: 111). In questo tipo di attività, il *post-editor* non mira a migliorare lo stile o la scorrevolezza della traduzione, bensì si limita a rendere il testo funzionale e comprensibile. Il *light post-editing* può essere equiparato a quello che Allen (2003: 302) definisce come *rapid post-editing*. Si tratta di un approccio impiegato per ottenere testi urgenti contenenti informazioni *time-sensitive*, destinate a scopo informativo e a una distribuzione ristretta, come verbali di incontri e report tecnici. Tale approccio è tipicamente adottato per l’*inbound translation*, cioè per traduzioni a uso interno²² con l’obiettivo di ottenere un testo in poco tempo ma con basse aspettative sulla qualità (Monti 2019: 111), con riferimento al concetto di *inbound translation*

²¹ Le competenze del *post-editor* saranno approfondite nel paragrafo 3.6.

²² Si tratta di testi non destinati alla pubblicazione, realizzati a scopo informativo, la cui traduzione mira a veicolare il senso generale del testo.

approach introdotto da Allen (2003: 301) che lo definisce come il processo di traduzione per la comprensione.

Il *full post-editing* è inteso come una revisione del testo tradotto automaticamente con l'intento di raggiungere una qualità pubblicabile (*outbound translation*). Il testo ottenuto dal *full post-editing* non dovrebbe differenziarsi da un testo prodotto da un traduttore professionista umano (Monti 2019: 111). Si parla in questo caso di *outbound translation approach*, concetto formulato da Allen (2003: 301), in cui l'*outbound translation* viene intesa come il processo di traduzione per la comunicazione verso l'esterno.

Il *full post-editing* si configura come un processo più lungo e più complesso rispetto al *light post-editing*. Ciò è dovuto all'obiettivo da raggiungere, ossia la massima accuratezza nella trasmissione del messaggio del testo di partenza mediante il testo di arrivo e la sua massima utilizzabilità (Monti 2019: 111).

Lo standard *ISO 18587* (2017), presentato più nel dettaglio nel paragrafo 3.4 del presente capitolo, fornisce le seguenti definizioni di *light post-editing* e *full post-editing*: il *light post-editing* è il "process of post-editing to obtain a merely comprehensible text without any attempt to produce a product comparable to a product obtained by human translation"; il *full post-editing* è il "process of post-editing to obtain a product comparable to a product obtained by human translation". La distinzione tra i due approcci riflette quindi non solo un diverso grado di intervento del *post-editor* sull'output della TA, ma anche esigenze e aspettative differenti in termini di destinazione d'uso e qualità finale del testo.

Il concetto di *full post-editing* solleva diversi interrogativi circa la sua reale efficienza: un testo sottoposto a questo tipo di intervento presenta un livello elevato di qualità; pertanto, è legittimo domandarsi se sia più conveniente post-editare l'output grezzo della TA o se sia preferibile procedere con una traduzione del documento ex novo (Allen 2003: 306). Infatti, il tempo e lo sforzo richiesti per un *full post-editing* potrebbero avvicinarsi a quelli necessari per una traduzione umana, riducendo così il vantaggio che la TA dovrebbe offrire in termini di efficienza. Prima di scegliere quale strategia traduttiva adoperare, se tradurre ex novo o post-editare, è necessario avere familiarità con il genere testuale e l'ambito specialistico trattati.

Per quanto riguarda la scelta del livello di *post-editing*, questa può essere dettata da diversi fattori. A tal proposito, Allen (2003: 301) ha individuato le seguenti variabili: l'utente/cliente; la quantità di documentazione da processare; le aspettative riguardanti il livello di qualità del prodotto finale; il tempo a disposizione; l'uso del documento considerando l'aspettativa di vita e la deperibilità delle informazioni; l'uso del testo finale. In definitiva, il livello di *post-editing* da

adottare va scelto accuratamente, in base al contesto e agli obiettivi comunicativi della traduzione, tenendo conto del delicato equilibrio tra tempi, costi e qualità.

3.2.1 Linee guida per il *light* e il *full post-editing*

In *TAUS MT Post-Editing Guidelines* (Massardo et al. 2016: 17), la qualità di un testo sottoposto a *light post-editing* è definita “good enough”. Massardo et al. (2016: 17) individuano i seguenti criteri per una qualità di questo tipo:

- il contenuto del testo deve essere comprensibile;
- il significato del testo di partenza deve essere preservato;
- la qualità stilistica assume un ruolo minoritario.

Alla luce delle sopramenzionate linee guida, gli autori individuano le seguenti indicazioni per l'esecuzione del *light post-editing*:

- ottenere una traduzione corretta dal punto di vista semantico;
- assicurarsi che nessuna informazione sia stata omessa o aggiunta accidentalmente;
- modificare i contenuti offensivi, inappropriati o culturalmente inaccettabili;
- utilizzare il più possibile l'output grezzo;
- applicare le regole ortografiche di base;
- non adoperare correzioni di natura esclusivamente stilistica;
- non modificare la struttura della frase per migliorarne la scorrevolezza.

Sulla base di quanto esposto, risulta chiaro che il prodotto di un *light post-editing* potrebbe essere un testo che non è idiomatico e scorrevole e in cui la grammatica e la sintassi non sono corrette. L'obiettivo principale, come citato precedentemente, resta la comprensibilità.

Per l'esecuzione del *full post-editing*, le *TAUS MT Post-Editing Guidelines* (Massardo et al. 2016: 18) individuano le seguenti linee guida:

- ottenere una traduzione corretta dal punto di vista grammaticale, sintattico e semantico;
- assicurarsi che la terminologia chiave sia tradotta correttamente e che i termini non tradotti appartengano alla lista fornita dal cliente di “termini da non tradurre”;
- assicurarsi che nessuna informazione sia stata omessa o aggiunta accidentalmente;
- modificare i contenuti offensivi, inappropriati o culturalmente inaccettabili;
- utilizzare il più possibile l'output grezzo;
- applicare le regole di base relative all'ortografia, alla punteggiatura e alla sillabazione;
- assicurarsi che la formattazione sia corretta.

Tenendo conto delle indicazioni sopra menzionate, emerge chiaramente che nel processo di *full post-editing* l'aspetto stilistico e la correttezza grammaticale e sintattica assumono un ruolo centrale, in quanto questo procedimento mira a raggiungere una qualità il più possibile vicina a quella di una traduzione realizzata da un traduttore professionista. Questi requisiti evidenziano le differenze tra la qualità posta come obiettivo dal *light post-editing* rispetto a quella alla quale tende il *full post-editing*. Il post-editor, in ogni caso, non deve dimenticare di utilizzare il più possibile l'output grezzo, come sottolineato nelle linee guida di entrambe le modalità di *post-editing*.

3.2.2 Necessità di aggiornamento delle linee guida per *light* e *full post-editing*

Diversi studi, tra cui quello di Rico Pérez (2024), mettono in evidenza come il confine tra gli approcci di *light post-editing* e *full post-editing* sia meno netto di quanto si supponga. In particolare, i traduttori possono riscontrare delle difficoltà nel distinguere una correzione essenziale da una correzione preferenziale (*preferential change*) e conseguentemente nel decidere se optare per un *full post-editing* o un *light post-editing* (O'Brien 2011: 19).

Come già evidenziato in precedenza, il concetto stesso di “intervento poco invasivo”, centrale nella definizione di *light post-editing*, è soggetto a interpretazioni variabili e, pertanto, risulta difficile da definire. Questo procedimento, oltre a essere influenzato da fattori come il pubblico destinatario e le esigenze del cliente, può essere condizionato anche dall'esperienza del *post-editor* e da aspetti psicologici individuali, come l'inclinazione al perfezionismo del traduttore che può tendere, quindi, ad apportare più modifiche di quelle realmente necessarie.

Inoltre, dallo studio condotto da Hu e Cadwell (2016) emerge che, sebbene la distinzione tra *light* e *full post-editing* sia ormai considerata uno standard nella pratica del *post-editing*, essa viene spesso adattata in modo informale sulla base delle esigenze del progetto, contribuendo a creare confusione tra i professionisti. Molti clienti, infatti, non conoscono le differenze tra i due approcci e i traduttori possono avere difficoltà nell'individuazione dell'approccio più consono da adottare per soddisfare le aspettative del cliente (Rico Pérez 2024: 29). In questo contesto, non sorprende che, come sostengono Koponen e Salmi (2018), i *post-editor* tendano a effettuare più correzioni del necessario, complice la poca chiarezza sull'implementazione pratica delle linee guida.

Partendo da queste criticità, Rico Pérez (2024) sottolinea la necessità di ripensare le linee guida di *post-editing* per riadattarle alla situazione concreta da affrontare, ai bisogni e alle aspettative dei traduttori da soddisfare, alla natura del compito, alle reali capacità del sistema di TA e ai tipi di errori attesi. Rico Pérez (2024: 34) parla di ecosistema di linee guida del *post-editing* (*MTPE guidelines ecosystem*), articolato in tre componenti fondamentali: informazioni

contestualizzate (*situated information*); il testo da post-editare (*the text*); le istruzioni per il *post-editing* (*MTPE instructions*). La Figura 6 mostra come si struttura questo ecosistema.

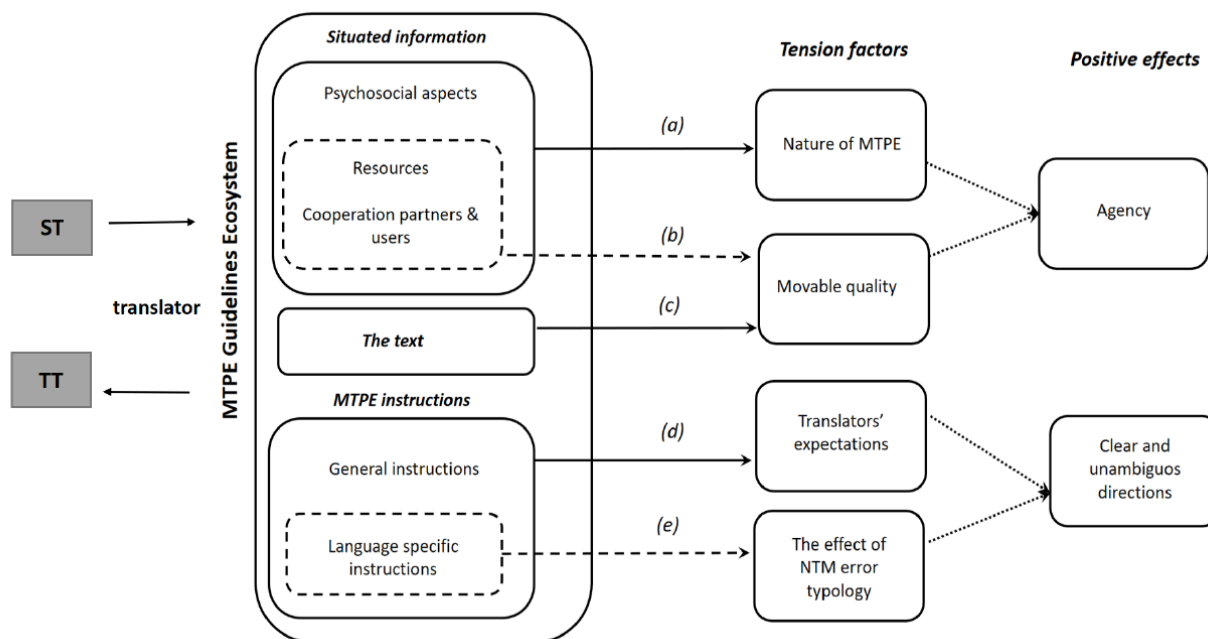


Figura 6: Ecosistema di linee guida del *post-editing*, Rico Pérez (2024: 34)

Rico Pérez (2024: 35-38) descrive dettagliatamente gli elementi costitutivi dell'ecosistema, come qui di seguito illustrato.

1. Informazioni contestualizzate: si riprende il concetto di modello situazionale di Krüger (2016b, 2016a) che pone il traduttore, visto come attore principale, al centro dell'ecosistema, di cui si individuano tre componenti essenziali:
 - aspetti psicosociali: si fa riferimento in particolare all'ambiente lavorativo e allo status professionale, includendo fattori come la motivazione, la pressione lavorativa, le tariffe, le informazioni sullo scopo e le tempistiche del progetto;
 - risorse: il traduttore deve essere a conoscenza delle risorse disponibili (sistemi di TA, *training data*, CAT tool, strumenti di allineamento, sistemi di memorie di traduzione e tecnologie come *file manager*, *checker*, ecc.);
 - partner e utenti: il traduttore deve disporre di informazioni sui suoi partner, sulle rispettive responsabilità e necessità, deve sapere con chi collabora, come organizzare il lavoro e quali sono le aspettative dell'utente finale.
2. Testo da post-editare: il testo viene analizzato in relazione al dominio specialistico, al rischio informativo, agli scopi comunicativi e al ciclo di vita del contenuto.

3. Le istruzioni per il *post-editing* vengono distinte in:

- istruzioni generali basate sul compito di *check and correct*, secondo cui l'autore confronta l'output della TA rispetto al testo di partenza (*check*) e apporta delle modifiche (*correct*);
- istruzioni specifiche sul linguaggio, ovvero orientate all'adeguamento lessicale e sintattico secondo le norme linguistiche del contesto di arrivo.

La Figura 6 mostra come i traduttori siano potenzialmente in grado di comprendere come eseguire il *post-editing* se dispongono di informazioni sul sistema di TA usato, sulla qualità dell'output attesa, sulle aspettative del cliente e sull'uso finale del testo.

In ultima analisi, come sostiene Rico Pérez (2024: 41), le linee guida strutturate secondo un approccio ecosistemico intendono contribuire non solo ad aumentare l'efficienza e la qualità del *post-editing*, ma anche a ridurre la resistenza dei traduttori verso la TA, rafforzandone la fiducia nel processo e migliorandone la motivazione e l'efficienza.

In questa ricerca si aderisce alla posizione espressa da Rico Pérez (2024) e sopra illustrata, riconoscendo che la considerazione dei fattori citati può risultare utile per l'elaborazione di linee guida mirate allo svolgimento di compiti traduttivi riguardanti specifici domini o tipologie di committenza. Inoltre, come affermano Hu e Cadwell (2016: 351), una discussione e una creazione congiunta da parte dei fornitori di servizi di traduzione e dei clienti di linee guida personalizzate per il *post-editing* potrebbe facilitare lo svolgimento del processo. I fattori sopramenzionati saranno tenuti in considerazione, ove possibile, nello sviluppo delle linee guida elaborate per il presente studio.

3.3. *Post-editing* monolingue vs *post-editing* bilingue

Nello scenario tradizionale, il *post-editing* è concepito come un'attività svolta da un professionista dotato di competenze linguistiche sia nella lingua di partenza che in quella di arrivo. Tuttavia, Pym (2014: 491) mette in discussione tale assunto, ipotizzando che la padronanza della lingua di partenza non sia sempre indispensabile per l'esecuzione del *post-editing*. Questa riflessione porta alla distinzione tra *post-editing* monolingue e *post-editing* bilingue.

Il *post-editing* bilingue implica un confronto diretto tra il testo di partenza e il testo generato dalla TA, con l'obiettivo di verificarne non solo la correttezza linguistica, ma anche l'adeguatezza semantica. Al contrario, il *post-editing* monolingue si concentra esclusivamente sull'output tradotto e garantisce la qualità del testo finale senza fare riferimento al testo di partenza (Nitzke e Hansen-Schirra 2021: 33).

Mitchell et al. (2013: 37) delineano delle linee guida per entrambi gli approcci. Per il *post-editing* monolingue, gli autori raccomandano di intervenire sul testo migliorandone la chiarezza e la scorrevolezza sulla base dell'interpretazione del significato, di correggere gli errori come l'ordine delle parole o l'ortografia nei casi in cui questi ostacolano la comprensione e di mantenere invariati quegli elementi linguistici che, pur non perfetti, risultino comunemente accettabili. Nel caso del *post-editing* bilingue, invece, il focus si sposta sulla verifica della correttezza semantica della traduzione rispetto al testo di partenza e sull'identificazione di omissioni o aggiunte improprie, sempre evitando modifiche non necessarie.

Nella pratica professionale, lo scenario più comune in cui viene richiesto al *post-editor* di eseguire un *post-editing* monolingue è quello in cui gli viene fornito l'output della TA con la richiesta di correggerlo e renderlo più scorrevole. Tuttavia, il *post-editor* può trovarsi più comunemente davanti a una richiesta di *post-editing* bilingue. Questo si rende necessario, ad esempio, quando l'output della TA appare incomprensibile e/o privo di senso oppure incoerente rispetto ad altre versioni tradotte dello stesso testo di partenza. In simili contesti, diventa difficile per il *post-editor* modificare l'output senza avere competenze nella lingua di partenza.

Come suggeriscono Nitzke e Hansen-Schirra (2021: 33), pur presentando il vantaggio di essere nettamente più veloce rispetto al *post-editing* bilingue (Krings 2001: 532), l'uso del *post-editing* monolingue andrebbe scoraggiato nell'ambito di una esecuzione professionale del *post-editing* in quanto nell'output del TA si possono verificare errori che non possono essere individuati se non si consulta il testo di partenza. Già Edmundson e Hayes (1985), due dei primi autori ad aver scritto sul *post-editing*, suggerirono questo tipo di approccio supportato da un codice grammaticale in grado di guidare il *post-editor* nel processo di correzione senza la necessità di consultare il testo di partenza. Tuttavia, come evidenziato da Krings (2001) mentre alcuni errori, ad esempio stilistici e/o grammaticali, sono facilmente rilevabili e di conseguenza correggibili senza servirsi del testo di partenza, altri errori, ad esempio di natura semantica, possono passare inosservati, provocando una distorsione del significato delle frasi. Pertanto, sebbene il *post-editing* monolingue possa risultare più efficiente in termini di tempo, esso comporta potenzialmente una perdita significativa di accuratezza, rendendolo una scelta meno affidabile in contesti in cui la qualità del contenuto e quindi l'accuratezza della trasmissione del messaggio sono cruciali.

Alla luce delle considerazioni espresse nel presente paragrafo, il *post-editing* nel presente elaborato è inteso esclusivamente come bilingue.

3.4 ISO 18587: *post-editing* standard

ISO 18587 è uno standard internazionale per il *post-editing*, la cui prima edizione è stata pubblicata nell'aprile del 2017. Lo standard non si applica agli sviluppi generali della TA o al processo generale di traduzione, bensì esclusivamente alla fase del *post-editing* e, più nello specifico, al *full post-editing* che viene definito come segue: “Post-editing is performed on MT output for the purpose of checking its accuracy and comprehensibility, improving the text, making the text more readable, and correcting errors.” (ISO 18587 2017).

Analogamente alle linee guida TAUS, lo standard ISO 18587 definisce i seguenti obiettivi per il processo di *post-editing*:

- l'output della TA sottoposto a *post-editing* deve essere comprensibile;
- il contenuto del testo di partenza deve corrispondere al contenuto della lingua d'arrivo;
- il *post-editor* deve rispettare i requisiti e le specifiche concordate.

La fase di *post-editing* coinvolge tre attori principali: il cliente, il fornitore di servizi di traduzione e il *post-editor*. Secondo lo standard, spetta al fornitore di servizi di traduzione stabilire l'idoneità del contenuto del testo di partenza alla TA e, di conseguenza, al *post-editing*. L'efficienza può dipendere dal tipo di sistema di TA impiegato, dalle lingue, dal dominio specialistico e dalla tipologia testuale. Il fornitore di servizi di traduzione deve inoltre decidere se è necessario il *pre-editing*. Il *post-editor* riceve, a sua volta, le informazioni specifiche relative al destinatario del testo di arrivo e alla qualità da raggiungere, con l'obbligo di rispettare i requisiti concordati dal cliente e dal fornitore.

Lo standard definisce il risultato del *post-editing* come un testo che non si distingue da una traduzione eseguita da un traduttore umano. Il testo deve essere accurato, comprensibile e stilisticamente accettabile: non deve contenere errori di grammatica, sintassi o punteggiatura; inoltre, deve presentare un uso coerente della terminologia. Analogamente a quanto stabilito dalle linee guida TAUS, il *post-editor* deve usare il più possibile l'output generato dalla TA, intervenendo solo sui contenuti non appropriati. Egli deve, infine, assicurarsi che le informazioni non siano omesse o aggiunte impropriamente. Per soddisfare tali standard qualitativi, la norma delinea un profilo professionale altamente specializzato, individuando sei competenze obbligatorie:

1. competenze traduttive;
2. competenze linguistiche nella lingua di partenza e di arrivo;
3. competenze testuali nella lingua di partenza e di arrivo;
4. competenze culturali per garantire la comprensione del testo sottoposto a *post-editing*;
5. competenze tecniche;

6. conoscenze del dominio specialistico trattato nel testo.

Lo standard definisce altri compiti che spettano al *post-editor* e al fornitore di servizi di traduzione. Il *post-editor* deve essere innanzitutto in grado di valutare se l'output della TA necessita di modifiche rispetto al testo di partenza e quindi di fornire un testo di arrivo impiegando gli elementi tradotti automaticamente o creando una nuova traduzione. A conclusione del processo di *post-editing*, il fornitore di servizi di traduzione controlla il testo finale e lo consegna al cliente. Il *post-editor* deve essere in grado di fornire un feedback sulle prestazioni del sistema di TA, così da consentire l'individuazione dei punti deboli e il miglioramento del sistema stesso.

In sintesi, lo standard ISO 18587 fornisce indicazioni sulle competenze del *post-editor* e del fornitore di servizi di traduzione e su come eseguire il *full post-editing* di testi tradotti da un sistema di TA.

3.5 Le competenze del *post-editor*

La figura professionale del *post-editor* si distingue per un insieme di competenze specifiche che vanno ben oltre la semplice correzione di testi generati automaticamente. Il pilastro centrale delle competenze del *post-editor* è la gestione degli errori, che implica non solo la capacità di individuare e classificare le anomalie presenti nell'output della TA, ma anche di valutarne la rilevanza e decidere se e come intervenire, in modo rapido ed efficace (Nitzke e Hansen-Schirra 2021: 74-75). Pertanto, il compito assegnato al *post-editor* comporta le seguenti abilità e competenze:

- distinguere il *post-editing* dalla revisione tradizionale, riconoscendo che gli errori da correggere sono di natura diversa e richiedono strategie specifiche;
- distinguere il *post-editing* dal lavoro in un ambiente CAT, modalità di lavoro che costituisce la norma per i traduttori professionisti, comprendendo le differenze operative e funzionali;
- applicare conoscenze acquisite prevalentemente attraverso la pratica professionale, poiché nella formazione universitaria si offre una preparazione limitata sul *post-editing*;
- possedere un'eccellente competenza digitale, con particolare riferimento alle abilità di videoscrittura ed elaborazione di testi;
- mostrare una buona resistenza e attitudine a lavorare direttamente sullo schermo del PC (eventualmente anche per periodi prolungati);
- prendere decisioni rapide e mirate in merito agli interventi di correzione, valutando di volta in volta il livello di accettabilità dell'output;
- bilanciare costantemente sforzo, tempo e costi, adeguando il proprio intervento agli obiettivi e alle richieste del progetto di traduzione.

Le caratteristiche sopra descritte sono strettamente legate all’attuazione pratica del task di *post-editing*. Torrejón e Rico (2012: 169-170) offrono una classificazione sistematica delle abilità richieste al *post-editor*, suddividendole in tre macrocategorie: *linguistic skills* (competenze linguistiche), *instrumental competence* (competenze strumentali) e *core competences* (competenze di base). Gli autori le strutturano nello schema seguente (cfr. Fig. 7).

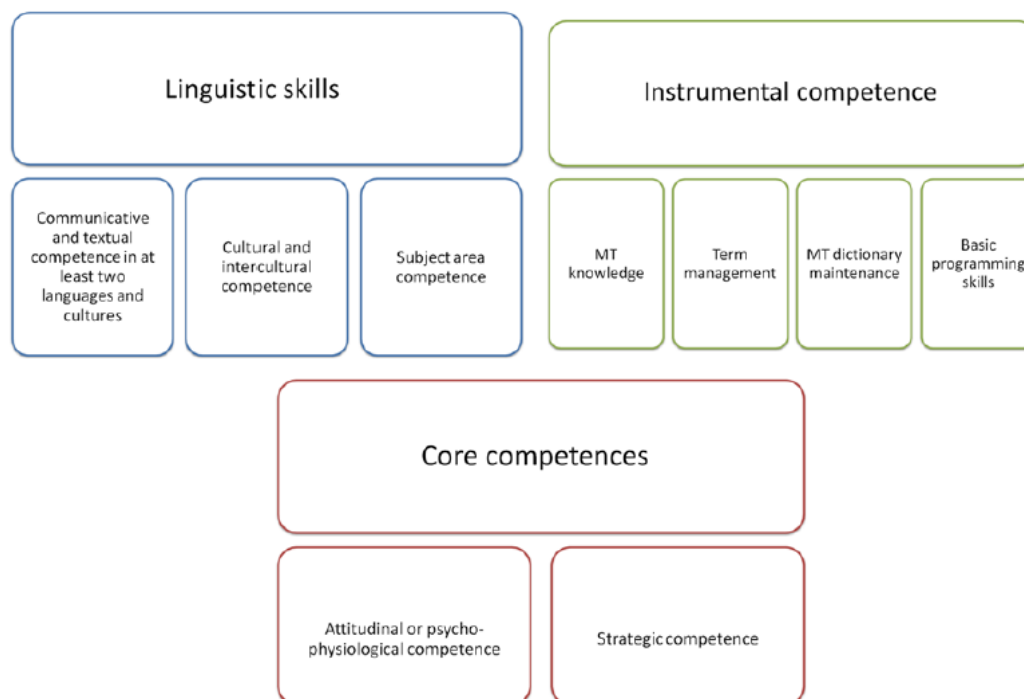


Figura 7: Abilità e competenze per il *post-editing*, Torrejón e Rico (2012: 169)

Torrejón e Rico (2012: 170) riconducono le competenze linguistiche necessarie per il *post-editing* a quelle proprie della figura professionale del traduttore. Sono quindi richieste competenze sia nella lingua di partenza sia in quella di arrivo, oltre a una solida conoscenza delle rispettive culture e dell’ambito tematico di riferimento. Anche nell’ambito delle competenze strumentali si riscontrano abilità tipiche del traduttore professionista, come ad esempio la capacità di gestire in modo efficace la terminologia. A ciò si aggiungono competenze riguardanti la TA, i suoi sistemi e le diverse architetture, come anche conoscenze di programmazione e *pre-editing* (Torrejón e Rico 2012: 170). Si tratta quindi di competenze che consentono al *post-editor* di utilizzare in modo funzionale le tecnologie coinvolte nello svolgimento del suo lavoro. Tali competenze strumentali possono essere ricondotte alle competenze tecniche previste dallo standard ISO 18587, richiamate nel paragrafo 3.4 del presente capitolo.

L’ultima macrocategoria teorizzata da Torrejón e Rico (2012) comprende le competenze di base, intese come un insieme di competenze attitudinali, psicofisiologiche e strategiche che permettono al *post-editor* di adottare l’approccio più adeguato all’output della TA a seconda del

compito da svolgere. Ciò implica una spiccata capacità decisionale: il professionista deve saper selezionare le norme da impiegare e applicarle, determinarne le modalità applicative, calibrare il livello di intervento necessario e, non da ultimo, gestire le aspettative del committente in relazione agli standard qualitativi attesi (Torrejón e Rico 2012: 170). Lo standard ISO 18587 (cfr. 3.4) non prende in considerazione quest'ultima macrocategoria nella lista delle competenze citate che, tuttavia, possono essere individuate come competenze alla base dei compiti assegnati al *post-editor*.

Le competenze e abilità evidenziate nello schema proposto da Torrejón e Rico (2012: 169-170) possono essere riassunte nella seguente lista:

- conoscenza generale della TA e dei suoi aspetti problematici;
- conoscenza dello specifico sistema di TA impiegato;
- conoscenza della lingua di partenza e di arrivo;
- familiarità con il dominio e la tipologia testuale;
- capacità di seguire le specifiche richieste per ogni lavoro.

In alcuni casi la conoscenza della lingua di partenza potrebbe non essere necessaria, in quanto, in linea teorica, una persona madrelingua competente in un dominio potrebbe essere in grado di individuare eventuali anomalie logiche e/o semantiche, prescindendo dal confronto con il testo di partenza. Ciò si ricollega al concetto di *post-editing* monolingue, trattato nelle sue peculiarità e nei suoi aspetti problematici nel paragrafo 3.3.

In merito alle competenze necessarie per il *post-editing*, Nitzke et al. (2019: 251) fanno accenno anche all'autopercezione e all'etica professionale: secondo gli studiosi, lo sviluppo della professione del *post-editor* dovrebbe essere accompagnato da una nuova consapevolezza secondo cui il *post-editor* non dovrebbe percepire sé stesso come un mero "correttore" dei testi prodotti da un sistema di TA, bensì come un vero e proprio esperto linguistico competente nell'esecuzione dei processi di *post-editing* e responsabile della qualità finale del testo.

Alla luce di quanto illustrato e come evidenziato da Nitzke et al. (2019: 251), va riconosciuto al *post-editor* il merito della creazione del testo d'arrivo, sebbene quest'ultimo non risulti sempre perfetto, considerando gli aspetti precedentemente esposti che ne caratterizzano il processo. In questo senso, il *post-editing* si configura come un'attività complessa e strategica, che richiede al *post-editor* competenze trasversali e una visione consapevole del proprio contributo all'interno del processo traduttivo.

3.6 La misurazione dell'*effort* e la retribuzione del *post-editor*

Nel contesto del *post-editing* della TA, la misurazione dell'*effort* (sforzo) richiesto rappresenta un tema centrale, sia per verificare l'effettiva efficienza del processo rispetto alla traduzione umana tradizionale, sia per definire criteri equi di retribuzione. Come dichiara Koponen (2016), a partire dai primi anni 2000, si è intensificata la ricerca sull'analisi del cosiddetto *post-editing effort*, termine che designa l'insieme delle risorse (materiali e non) impiegate dai *post-editor* per modificare l'output della TA e fare in modo che raggiunga la qualità attesa. Krings (2001) propone le seguenti tipologie di *post-editing effort*, ormai ampiamente consolidate:

- *temporal effort*;
- *technical effort*;
- *cognitive effort*.

Per *temporal effort* (sforzo temporale) si intende il tempo impiegato per lo svolgimento del *post-editing* e misurato, ad esempio, tenendo conto del numero di parole post-editate per secondo (O'Brien 2011: 204). In altre parole, questo tipo di *effort* tiene conto del tempo necessario per raggiungere la qualità desiderata. Tale criterio è spesso utilizzato come parametro principale in ambito commerciale, grazie alla sua facile misurabilità (O'Brien 2022: 115).

Il *technical effort* (sforzo tecnico), invece, si riferisce al numero e alla tipologia di modifiche manuali richieste sull'output della TA per raggiungere la qualità desiderata. In questo caso si parla di sforzo manuale (*manual effort*) con cui si intende in particolare:

- l'apportare delle correzioni (su carta o su supporto elettronico);
- lo "sforzo da tastiera", cioè le azioni compiute schiacciando i tasti della tastiera, come ad esempio cancella, aggiungi, sostituisci ecc.

Il *technical effort* può essere, quindi, quantificato tramite il numero di battute su tastiera, inserimenti, cancellazioni, sostituzioni, o mediante metriche di *edit distance*²³, utili a calcolare il numero minimo di operazioni necessarie per trasformare il testo generato dalla macchina in quello post-editato (Jia et al. 2019; O'Brien 2022: 116).

Tuttavia, entrambi questi tipi di sforzo non riescono a catturare appieno la dimensione più complessa del lavoro del *post-editor*, ovvero il *cognitive effort* (sforzo cognitivo). Krings (2001:182) lo definisce come: "the type and extent of cognitive processes triggered by the post-editing task". Questo tipo di *effort* si configura come il più strutturato, in quanto sottende diverse azioni da parte del *post-editor*: leggere e comprendere il testo; confrontare il significato del testo di

²³ Il concetto di *edit-distance* è stato già trattato nel paragrafo 2.2.2.1.

partenza con quello dell'output della TA; individuare e classificare gli errori; trovare e formulare una soluzione adeguata. Lo sforzo cognitivo può essere calcolato solo in modo indiretto. Tra gli strumenti che possono essere impiegati, O'Brien (2022: 117) individua i seguenti:

- *Think aloud protocols*: il *post-editor* riferisce a voce alta sul suo lavoro e questo permette di evidenziare le difficoltà cognitive che può incontrare. Questo procedimento potrebbe tuttavia interferire con il suo lavoro e rallentarlo.
- Registrazioni del monitor mentre il *post-editor* lavora; successivamente, a conclusione del lavoro, si osserva dove sono state apportate le modifiche al testo e si chiede al *post-editor* di commentare eventuali difficoltà riscontrate. Lo svantaggio di questo strumento è che il *post-editor* potrebbe non ricordare tutti i problemi affrontati.
- *Eye-tracking*: si tratta di una tecnologia che registra i movimenti degli occhi, il tempo dedicato dall'occhio a una parte del testo (*fixation duration*) e la dilazione della pupilla. Questa tecnologia, unita alla registrazione dello schermo, è utile per osservare dove vengono apportate le modifiche. Le problematiche di questo metodo sono diverse: lo strumento per l'*eye-tracking* è estremamente costoso e l'interpretazione e la raccolta dei dati sono processi complessi. Inoltre, per lo svolgimento della rilevazione è necessario un ambiente controllato.

Un ulteriore metodo per misurare lo sforzo cognitivo è l'analisi delle pause tra le battute di tastiera, considerate come indicatori significativi dello sforzo cognitivo già in diversi studi sulla traduzione (cfr. ad es. Schilperoord 1996). Krings (2001) è stato tra i precursori nell'applicazione di questo metodo al *post-editing*, seguito poi da O'Brien (2006). Quest'ultima studiosa in particolare ha indagato la correlazione tra la traducibilità del testo di partenza mediante la TA e lo sforzo cognitivo del *post-editor*, riconoscendo all'analisi delle pause un ruolo fondamentale. Più recentemente, Jia et al. (2019) hanno confermato che pause più lunghe e più frequenti sono un segnale di uno sforzo cognitivo più elevato.

I diversi tipi di *effort* illustrati possono essere impiegati per calcolare la retribuzione del *post-editor*. A tal riguardo, esistono diversi modelli di pagamento:

- a seconda del numero delle parole del testo di partenza/testo di arrivo;
- a seconda delle ore impiegate;
- a seconda delle correzioni (*edit-distance*).

La determinazione della retribuzione basata esclusivamente sul volume delle correzioni apportate può condurre alla nascita di un conflitto di interessi, in quanto il *post-editor* potrebbe effettuare modifiche non necessarie per ottenere un compenso più elevato. Un'ulteriore criticità

riguardante l'uso di questa tipologia di pagamento è che considera soltanto lo sforzo manuale, sottovalutando conseguentemente il carico mentale e il valore decisionale del lavoro svolto. In questo modello, lo sforzo cognitivo risulta fortemente penalizzato, ignorando il fatto che l'attività di battitura su tastiera è solo la fase terminale e meno complessa dell'intero processo di *post-editing*.

Inoltre, l'entità dello sforzo richiesto può variare notevolmente in base a diversi fattori, tra cui la qualità dell'output della TA, la combinazione linguistica, il dominio del testo e l'esperienza del *post-editor* (Lacruz et al. 2012; O'Brien 2022). Infine, la misurazione dello sforzo, oltre a fornire indicazioni utili per la retribuzione, può essere interpretata come una misura indiretta della qualità del sistema di TA: maggiori sono i cambiamenti, il tempo e lo sforzo richiesto, minore è la qualità dell'output generato (O'Brien 2022: 117). In quest'ottica, il *post-editing effort* non è solo un parametro operativo, ma anche uno strumento valutativo nel più ampio processo di sviluppo e ottimizzazione dei sistemi di TA.

In conclusione, si può rilevare che la questione della retribuzione del lavoro del *post-editor* rimane tuttora irrisolta.

3.7 Linee guida per la classificazione degli errori

In letteratura sono state definite svariate linee guida utili alla categorizzazione sia degli errori della TA, sia di quelli umani. Le linee guida sviluppate da Flanagan (1994) sono tra le prime a essere state create per la classificazione degli errori della TA. Esse individuano categorie applicabili a un numero ampio di lingue (come, ad esempio, errori riguardanti l'uso delle preposizioni o dell'elisione) e categorie che si applicano esclusivamente a date coppie linguistiche, per esempio della coppia linguistica inglese-tedesco che presenta la problematica relativa alla resa del caso.

Un ulteriore contributo alla classificazione degli errori nella TA è stato offerto da Vilar et al. (2006), le cui linee guida si applicano in particolare a tre coppie linguistiche: spagnolo-inglese, inglese-spagnolo e cinese-inglese. Per quanto riguarda le prime due coppie linguistiche, il corpus sul quale si basa l'individuazione degli errori è costituito da testi di discorsi tenuti in sessioni plenarie del Parlamento europeo. Per la terza coppia linguistica, sono stati selezionati testi presenti nel *Linguistic Data Consortium* (LDC). Queste linee guida si compongono di cinque macrocategorie di errore (*missing words*, *word order*, *incorrect word*, *unknown words* e *punctuation*) che a loro volta presentano uno o due livelli di sottocategorie. Per la coppia linguistica cinese-inglese sono state introdotte categorie di errori aggiuntive, in particolare in riferimento alla macrocategoria *word order*.

Farrús et al. (2011) e Costa et al. (2015) propongono tassonomie che classificano gli errori di TA sulla base di diversi livelli linguistici. Le linee guida proposte da Farrús et al. (2011) si basano

su un corpus parallelo spagnolo-catalano composto da frasi tratte dal quotidiano *El Periodico*. Si prendono in esame diversi livelli linguistici (ortografia, morfologia, lessico, semantica e sintassi) e per ognuno si individuano le tipologie di errore. Invece, le linee guida elaborate da Costa et al. (2015) prendono in esame la coppia linguistica inglese-portoghese, individuando gli errori su cinque diversi livelli linguistici (ortografia, lessico, semantica, grammatica e dialogo), per ognuno dei quali si definiscono categorie di errori, che, a loro volta, possono presentare diverse sottocategorie. Il corpus su cui si basano queste linee guida si compone di trascrizioni e traduzioni di discorsi (TED-talks), traduzioni nell'ambito di un dominio specialistico (testi turistici) e traduzioni di costruzioni specifiche (domande).

È possibile annoverare tra le linee guida per la classificazione di errori umani e della TA le seguenti tassonomie: *SAE J2450* e *LISA QA Modell*. Entrambe forniscono un elenco di tipi di errore e li classificano in base alla gravità con l'intento di calcolare i punteggi di qualità. *SAE J2450* nasce nel 1997 in seguito a una task force realizzata da SAE (Society of Automotive Engineers) con l'intento di definire uno standard di traduzione e verificare la qualità della resa di informazioni sul settore automobilistico (Schütz 1999: 278). *LISA QA Modell*, invece, è stata sviluppata nel 1995 e distribuita da LISA (Localization Industry Standards Association) per i progetti di localizzazione. La sua ultima versione (versione 3.1) è stata pubblicata nel 2004.

Le valutazioni della qualità della traduzione a livello di frase sono altamente variabili poiché legate alla soggettività del *post-editor*. Con l'intento di limitare questa variabilità, White (1995) presenta una distinzione tra *adequacy*, *fluency* e *informativeness*. Mentre l'*adequacy* riguarda la quantità di contenuto e significato del testo di partenza e come questa viene trasmessa nel testo di arrivo, la *fluency* riguarda la misura in cui la traduzione presenta una buona scorrevolezza, indipendentemente dal significato della frase. L'*informativeness*, invece, riguarda la capacità del testo di arrivo di trasmettere le informazioni necessarie al destinatario.

Le distinzioni presentate da White (1995) sono state successivamente riprese da Lommel, Uszkoreit, et al. (2014), i quali hanno introdotto più nello specifico la distinzione tra *accuracy* e *fluency* per la classificazione di errori di traduzione nella tassonomia *Multidimensional Quality Metrics* (MQM). Tali linee guida sono state sviluppate nell'ambito dei progetti di ricerca QTLaunchPad e QT21, finanziati dall'Unione Europea in collaborazione con il Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI). Esse si compongono di più di 100 categorie utili alla classificazione di errori umani e della TA e hanno l'aspirazione di adattarsi a qualsiasi genere testuale e ambito specialistico²⁴. La prima versione²⁵, lanciata nel 2015, è stata gradualmente

²⁴ Per ulteriori informazioni, consultare il sito: <https://themqm.org/> (ultima consultazione: 20/12/2025)

sostituita da versione sviluppate successivamente, la cui più recente presenta sette livelli superiori di tipi di errore, che a loro volta presentano dei sottolivelli²⁶. In quest'ultima versione la macrocategoria di *fluency* è stata sostituita dal livello *linguistic conventions*.

La distinzione tra *accuracy* e *fluency* è stata impiegata anche da Tezcan et al. (2017) nella *SCATE (Smart Computer-aided Translation Environment) MT error taxonomy*. Queste linee guida si applicano agli errori della TA riguardo alla coppia linguistica inglese-olandese. Sulla base di *SCATE MT error taxonomy* si è sviluppato un corpus, *SCATE corpus of MT errors*, che si compone di oltre 3600 coppie di frasi annotate. Esso si differenzia dai corpora annotati già esistenti in quanto gli errori di *accuracy* sono annotati creando un collegamento tra il segmento dell'output contenente l'errore e il corrispondente segmento del testo di partenza. Questo permette di determinare i frammenti del testo di partenza problematici per un sistema di TA.

Sebbene la maggior parte delle tassonomie descritte nel presente paragrafo fornisca criteri sistematici per l'identificazione delle tipologie d'errore, esse non informano gli annotatori sugli intervalli di annotazione, che potrebbero essere variabili, né includono esempi. Un'eccezione è rappresentata dal lavoro di Lommel, Uszkoreit, et al. (2014), che contiene linee guida dettagliate su come e dove gli errori debbano essere annotati.

A conclusione di questa rassegna riguardante le linee guida per la classificazione degli errori, preme sottolineare che la selezione operata non ha pretese di esaustività; l'obiettivo del presente paragrafo è piuttosto delineare un quadro teorico dei principali standard di riferimento per la classificazione degli errori, siano essi di matrice umana o generati dai sistemi di TA.

3.8 Sfide e aspetti problematici del *post-editing*

Il *post-editing* si configura come una pratica tanto promettente quanto complessa, riguardo alla quale emergono numerose sfide, sia teoriche che operative. Una delle difficoltà principali riguarda la definizione stessa di qualità. Come osserva Rico Pérez (2024: 32), i traduttori e i revisori devono stabilire cosa significhi “qualità” per decidere se intervenire o meno su un segmento generato automaticamente. Tuttavia, in questo contesto la qualità non è un valore assoluto: non si parla di una traduzione corretta o errata, bensì di una resa “sufficientemente accettabile per uno scopo specifico” (Rico Pérez 2024: 21). Tale considerazione si collega alla visione di Way (2013) secondo cui la soglia di accettabilità dipende dal ciclo di vita del documento: un testo con contenuti funzionali a un breve utilizzo, come il verbale di un incontro, potrebbe richiedere solo un controllo minimo, mentre

²⁵ È possibile consultare la prima versione al seguente link:

<https://web.archive.org/web/20211216145215/https://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html> (ultima consultazione: 20/12/2025)

²⁶ Per ulteriori informazioni sulla versione più recente di MQM, consultare il sito: <https://themqm.org/the-mqm-full-typology/> (ultima consultazione: 20/12/2025)

contenuti destinati a un uso più duraturo, ad esempio della documentazione legale, necessitano di un intervento più approfondito.

A tal proposito, Nitzke et al. (2019: 246) propongono delle linee guida per aiutare il *post-editor* a decidere se usare la TA e quale tipo di *post-editing* effettuare. Gli autori suggeriscono l'impiego della TA seguita da un *light post-editing* solo per testi con una qualità attesa del 60-80% e un basso livello di rischio. Viceversa, per testi ad alta criticità e qualità richiesta superiore all'80%, l'uso della TA è sconsigliato o si suggerisce tutt'al più un *full post-editing*. Ciò implica un approccio "fit for purpose" (Bowker 2019; Way 2018), più focalizzato sull'adeguatezza all'uso finale che sulla fedeltà a una traduzione umana.

In considerazione di quanto esposto, si delinea la questione centrale dell'efficienza, ossia la determinazione della maggiore convenienza tra l'impiego del *post-editing* e l'esecuzione di una traduzione ex novo. Questa domanda nasce soprattutto in riferimento al *full post-editing*, processo che comporta come risultato finale la creazione di un testo la cui qualità sia comparabile a quella di un testo tradotto interamente dall'uomo. Diversi studi indicano che il *post-editing* sia, nella maggior parte dei casi, più rapido (O'Brien 2005; Plitt e Masselot 2010), ma non sempre ciò si traduce in un vantaggio netto, come testimonia lo studio di Screen (2017) in riferimento a testi che trattano un linguaggio comune. Inoltre, la produttività non va intesa solo in termini quantitativi, qualitativi e temporali: essa dipende anche dallo sforzo cognitivo richiesto. Quando tale sforzo è elevato, la produttività percepita cala sensibilmente (O'Brien 2011: 198).

Un ulteriore aspetto potenzialmente problematico riguarda la percezione soggettiva dei traduttori nei confronti del *post-editing*. Diversi studi mostrano che, pur essendo oggettivamente più produttivo (Gaspari et al. 2014: 70), il *post-editing* è vissuto come un compito meno stimolante (Moorkens e O'Brien 2017: 126) e spesso viene preferito meno della traduzione ex novo (Teixeira 2014: 56; Gaspari et al. 2014: 70). Secondo Alvarez-Vidal et al. (2020: 65), la ripetitività, l'imprevedibilità degli errori della TA, la conseguente difficoltà nell'individuarli e i vincoli creativi sono tra i fattori che influenzano negativamente l'esperienza dei *post-editor*. Gli autori rivelano, inoltre, che anche i traduttori più esperti lo considerano più faticoso, meno gratificante e meno retribuito, sebbene riconoscano che la produttività migliori con l'esperienza e con una formazione adeguata.

Pertanto, la formazione e l'esperienza sono elementi chiave per un *post-editing* funzionale, tuttavia dal punto di vista della formazione, la TA e il *post-editing* stanno tuttora entrando gradualmente a far parte dei programmi di laurea in traduzione come materie indipendenti (Guerrero Romeo 2018: 89). Con particolare riferimento al panorama della formazione universitaria

italiana, lo studio di Martín-Mor (2018) ha evidenziato che non tutti i corsi di laurea magistrale in traduzione includono corsi di tecnologie per la traduzione.

Un forte impulso è stato dato, e continua a essere dato, dall'*European Master's in Translation* (EMT)²⁷, che ha portato l'attenzione delle istituzioni accademiche su corsi esclusivamente dedicati alle tecnologie per la traduzione. Tra questi è possibile individuare il *MSc in Translation Technology* presso la Dublin City University²⁸ e il *Master Technologies de la traduction* presso l'Université de Lorraine²⁹. Nell'ambito dell'EMT è stato sviluppato l'*EMT Competence Framework*: si tratta di un quadro di riferimento riguardante le competenze necessarie per un traduttore e che devono essere sviluppate dagli studenti che frequentano questi corsi. Tra le competenze individuate figura la competenza tecnologica comprendente la capacità di implementare in modo critico la TA nel processo traduttivo³⁰. Tuttavia, la mancanza di figure educative esperte rappresenta ancora una sfida significativa per l'integrazione piena del *post-editing* nella didattica. Come sottolinea Gene (2021: 194), la sfida più urgente per le istituzioni formative è la carenza di docenti adeguatamente preparati a insegnare questa disciplina. Questo conferma ulteriormente quanto già menzionato circa l'introduzione del *post-editing* nei curricula universitari come un fenomeno relativamente recente.

Infine, si può affermare che il *post-editing* impone un oneroso carico cognitivo e richiede elevati livelli di concentrazione estremamente prolungati. Pause ergonomiche, in cui il *post-editor* è invitato a svolgere altre attività e a orientare i suoi pensieri verso idee differenti, unitamente a cambiamenti di ambiente, si rivelano fondamentali per mitigare l'affaticamento e preservare la costanza della performance nel tempo.

In conclusione, il *post-editing* è un'attività a basso margine creativo, ma ad alta intensità cognitiva, la cui efficacia dipende da variabili molteplici: dalla qualità della TA, alla competenza del traduttore, alla sua motivazione e al contesto professionale e formativo in cui opera. Altri fattori che possono influire sull'efficacia del *post-editing* riguardano le caratteristiche del testo da tradurre: la coppia linguistica, il linguaggio, il genere testuale e il dominio presi in esame. Nel capitolo successivo verrà trattato il linguaggio scientifico, oggetto di studio del presente progetto di ricerca.

²⁷ Un progetto di partenariato tra la Commissione Europea e gli istituti di istruzione superiore presente nel territorio europeo che offrono programmi di traduzione a livello di master. Per maggiori informazioni, consultare il link: https://commission.europa.eu/education/skills-and-qualifications/develop-your-language-skills/european-masters-translation-emt_en (ultima consultazione: 20/12/2025)

²⁸ Maggiori informazioni sono disponibili al seguente link: <https://www.dcu.ie/courses/postgraduate/salis/msc-translation-technology> (ultima consultazione: 20/12/2025)

²⁹ Maggiori informazioni sono disponibili al seguente link: <https://all-metz.univ-lorraine.fr/formation/langues-etranangeres-appliquees/master-technologies-de-la-traduction-tetra> (ultima consultazione: 20/12/2025)

³⁰ L'EMT competence framework è consultabile al seguente link: https://commission.europa.eu/document/download/b482a2c0-42df-4291-8bf8-923922ddc6e1_en?filename=emt_competence_fw_k_2022_en.pdf (ultima consultazione: 20/12/2025)

La scrittura scientifica

Nel presente capitolo vengono affrontati gli aspetti fondamentali della scrittura scientifica. In apertura si offre una panoramica delle principali tipologie di comunicazione scientifica, seguita dalla presentazione delle principali forme di testo scientifico con un approfondimento sugli elementi strutturali comuni. Il capitolo prosegue trattando nel dettaglio il linguaggio scientifico e le sue componenti fondamentali, per poi concentrarsi in modo specifico sulle peculiarità del linguaggio scientifico in tedesco e in italiano.

4.1 I livelli di comunicazione nella scienza

La comunicazione scientifica si manifesta in numerose e svariate forme. Graefen (1997: 9) riconosce come primo criterio di differenziazione l'area di studio specialistico, che si riflette principalmente sul contenuto e sulla terminologia specialistica impiegata. Ulteriori criteri di differenziazione riguardano i soggetti coinvolti e il genere testuale impiegato, sulla base dei quali è possibile individuare diversi livelli di comunicazione scientifica.

La prima categorizzazione presa in esame è quella proposta da Calaresu (2006: 40) che distingue tra comunicazione scientifica primaria e comunicazione scientifica secondaria. La prima tipologia si riferisce alla comunicazione tra specialisti; la seconda riguarda lo scambio comunicativo tra semi-specialisti, come gli insegnanti, e/o destinatari non specialisti, quali studenti o pubblico di massa). Stahlheber (1992: 171) distingue la comunicazione scientifica primaria, rappresentata da periodici specialisti destinati a esperti del settore, dalla comunicazione scientifica secondaria, che comprende invece le riviste orientate alla divulgazione delle conoscenze a un pubblico più ampio.

Un modello più articolato di comunicazione scientifica è quello presentato da Cloître e Shinn (1985), citato da Bucchi (2000: 143), che si struttura su quattro livelli:

1. *livello intraspecialistico* che riguarda la comunicazione tra esperti appartenenti allo stesso ambito disciplinare (ad esempio articoli scientifici nelle scienze fisiche e naturali, monografie nelle scienze umane e sociali);
2. *livello interspecialistico* che concerne lo scambio comunicativo tra studiosi di settori o sottosettori differenti (come nel caso di riviste scientifiche a carattere multidisciplinare);
3. *livello pedagogico* inerente alla trasmissione di conoscenze in ambito didattico (ad esempio libri di testo);

4. *livello popolare* volto a diffondere conoscenze presso un pubblico non esperto; in questo caso si parla anche di livello divulgativo (ad esempio stampa quotidiana e periodica, radio e televisione).

I livelli teorizzati da Cloître e Shinn riguardano la comunicazione scientifica non solo in forma scritta ma anche orale: ad esempio a livello intraspecialistico e interspecialistico s'intendono discussioni a convegni, a livello pedagogico lezioni frontali, a livello popolare programmi televisivi di divulgazione scientifica.

Loffler-Laurian (1983: 10-12) propone una categorizzazione del discorso scientifico articolata in sei tipi, individuati sulla base della situazione comunicativa, dei soggetti coinvolti e della natura del supporto mediante il quale il messaggio viene trasmesso:

1. discorso scientifico specialistico;
2. discorso di semi-divulgazione scientifica;
3. discorso di divulgazione scientifica;
4. discorso scientifico pedagogico;
5. discorso scientifico del tipo memoria, tesi, ecc.;
6. discorso scientifico ufficiale.

Il discorso scientifico specialistico riguarda la comunicazione prodotta da esperti che si rivolgono a colleghi dello stesso ambito professionale mediante pubblicazioni altamente specializzate. La comunicazione di tipo semi-divulgativo, invece, è realizzata ad esempio da ricercatori o giornalisti ed è destinata a un pubblico con conoscenze approfondite nell'ambito disciplinare di riferimento. Il discorso di divulgazione scientifica è prodotto, ad esempio, da giornalisti e si indirizza al grande pubblico su riviste pluridisciplinari o su riviste dedicate a un unico argomento con scopi informativi. Il gruppo comprendente il discorso scientifico pedagogico riguarda la comunicazione che si svolge tra docente e apprendente o tra giornalisti e un pubblico che si approccia all'argomento. Quello di tipo memoria/tesi, invece, riguarda studenti che approfondiscono una tema aspecifico o di specialisti che fanno il punto della situazione, rivolgendosi in entrambi i casi a una commissione di esperti. Infine, il discorso scientifico ufficiale è elaborato, ad esempio, da un ricercatore o un'équipe per un'istituzione o un'amministrazione e può presentarsi sotto forma di rapporti di ricerca, relazioni tecniche o documenti amministrativi standardizzati, come moduli o protocolli ufficiali.

Le prime quattro tipologie di discorso scientifico individuate da Loffler-Laurian (1983) possono essere ricondotte ai livelli di comunicazione individuati da Cloître e Shinn (1985), dove il discorso scientifico specialistico corrisponde al livello intraspecialistico, il discorso di semi-divulgazione

scientifica al livello interspecialistico, il discorso di divulgazione scientifica al livello popolare e il discorso scientifico pedagogico al livello pedagogico. Le ultime due forme, memoria/tesi e discorso ufficiale, non hanno invece un equivalente diretto nella classificazione di Cloître e Shinn. La prima rappresenta un genere ibrido: mantiene il rigore del livello intraspecialistico ma è prodotto in un contesto formativo e valutativo, non di scambio tra pari. Il discorso ufficiale, invece, ha una funzione prevalentemente amministrativa e istituzionale, con un pubblico variabile (specialistico o misto) e un formato fortemente standardizzato; per questo può essere considerato una categoria funzionale autonoma, non riconducibile a un singolo livello della tassonomia di Cloître e Shinn. Gli ultimi due livelli proposti da Löffler-Laurian (memoria/tesi e discorso ufficiale) appaiono meno rilevanti ai fini del presente studio. Per questi motivi, in questa sede si adotta come riferimento la classificazione di Cloître e Shinn.

Sulla base di quanto esposto sopra, si può concludere che la differenziazione principale riguarda emittenti e destinatari. L'autore del testo scientifico adatta il messaggio al pubblico di riferimento, determinando così il livello di comunicazione scientifica e influenzando la funzione del testo e le sue caratteristiche formali e stilistiche.

Il presente elaborato prende in esame una comunicazione tra esperti, in quanto vengono analizzati, articoli pubblicati su riviste scientifiche specialistiche, che riguardano aree tematiche specifiche e si indirizzano a un pubblico esperto. Riprendendo la distinzione di Cloître e Shinn (1985), si tratta di un livello di comunicazione intraspecialistico che dà vita a un discorso scientifico specialistico.

4.2 Le forme del testo scientifico scritto

Il testo scientifico assume prevalentemente due formati: libro e articolo. Luccio (2014: 5-11) offre una panoramica delle caratteristiche più rilevanti, di seguito riportate, dei libri scientifici. Questi si compongono di fogli di dimensioni standard, cioè pagine, rilegati. Il testo vero e proprio è preceduto da prefazione, premessa e indice. La prefazione costituisce un commento al testo realizzato da un autore diverso rispetto all'autore del testo del libro. Alla prefazione segue la premessa, in cui l'autore fornisce indicazioni di carattere generale: spiega ad esempio com'è nato il testo e rivolge ringraziamenti a persone e/o istituzioni che l'hanno supportato nella stesura. L'indice riporta i titoli dei capitoli, paragrafi e sottoparagrafi. Il testo si apre con l'introduzione, che inizia pertanto il lettore ai temi trattati nel testo: l'autore spiega le motivazioni che l'hanno condotto alla stesura del testo e illustra il contesto nel quale esso si colloca.

Come sottolinea Luccio (2014: 8-10), il testo di un libro si compone di capitoli che, a loro volta, possono essere raggruppati in parti sulla base di criteri comuni. Le parti, così come i capitoli,

presentano un titolo e un numero. Inoltre, di frequente i capitoli si suddividono in paragrafi, anch'essi dotati di titolo e numerazione progressiva, che a loro volta possono essere suddivisi in sottoparagrafi, chiamati anche paragrafi di secondo livello, e, in alcuni casi, anche in paragrafi di terzo livello.

Luccio (2014: 10-11) aggiunge che in alcuni libri scientifici è possibile trovare anche appendici e un glossario. L'appendice è una sezione posta alla fine del libro, che include informazioni supplementari che l'autore ritiene possano essere di interesse per il lettore, ma che sono marginali o troppo dettagliate per essere inserite nel testo vero e proprio. L'appendice è spesso seguita dal glossario, che costituisce una lista di termini poco familiari al lettore, accompagnati da una definizione. Alla fine della maggior parte dei libri è possibile trovare due indici analitici: uno dei nomi delle persone citate e uno degli argomenti. Se presenti entrambi, l'indice dei nomi precede l'indice degli argomenti.

Il secondo formato preso in considerazione è l'articolo scientifico, che Graefen (1997: 58) descrive come il risultato di un confronto con un tema scientifico di attualità o con un problema noto ma non ancora risolto. Si tratta di un contributo volto a favorire il progresso nell'ambito disciplinare di riferimento. Due caratteristiche fondamentali degli articoli scientifici meritano di essere evidenziate: da un lato, l'articolo è un testo autonomo, leggibile e comprensibile senza materiali aggiuntivi; dall'altro, esso rappresenta un contributo a un insieme più ampio di pubblicazioni, che deve pertanto subordinarsi allo scopo complessivo dell'insieme a cui appartiene (Graefen 1997: 60). Di frequente i testi scientifici sono pubblicati come articoli in riviste del settore, caratterizzate da periodicità variabile: settimanali, mensili, trimestrali, quadrimestrali, semestrali. Ogni rivista è contrassegnata da un numero che indica l'anno, il volume e il numero del volume (Luccio 2014: 12-13). I contributi scientifici possono essere pubblicati anche all'interno di volumi e monografie.

In riferimento alla struttura degli articoli scientifici, Luccio (2014: 13) identifica le principali componenti ricorrenti, che verranno approfondite in 4.2.2: 1) il titolo; 2) le parole chiave; 3) l'abstract; 4) l'introduzione; 5) il problema; 6) i materiali e i metodi; 7) i risultati; 8) la discussione; 9) le conclusioni; 10) la bibliografia. Le parti 5) e 9) possono essere eventualmente accorpate ad altre componenti, dando origine alla struttura consolidata degli articoli scientifici nota con l'acronimo IMRAD (*Introduction, Methods, Results, and Discussion*).

Allo stato attuale gli articoli scientifici sono pubblicati in formato stampato e in formato elettronico (PDF e HTML). La diffusione del supporto elettronico ha modificato ampiamente il panorama tradizionale della letteratura scientifica, favorendone l'accesso, ampliando la visibilità dei

contenuti e aprendo la strada a nuove modalità di consultazione, condivisione e fruizione dei risultati di ricerca.

4.2.1 Tipologie di articolo scientifico

Secondo Graefen (1997: 57) l'articolo scientifico nasce dalla consapevolezza dell'autore di possedere un vantaggio conoscitivo rispetto al pubblico, accompagnata dall'obbligo morale di condividere tale conoscenza a favore dello sviluppo del patrimonio scientifico collettivo. Data la vastità dei possibili argomenti trattati, l'articolo può assumere diverse tipologie.

Luccio (2014: 16-20) individua i principali tipi di articoli e le rispettive caratteristiche, a cui si fa riferimento di seguito.

- a) *Ricerche empiriche*: si tratta di *reports*, resoconti, in cui vengono presentate analisi di dati raccolti durante ricerche effettuate in laboratorio o sul campo. Tale tipologia di articoli si compone normalmente di paragrafi che presentano il seguente ordine: introduzione, metodo, risultati e discussione. In alcuni casi è possibile trovare anche un paragrafo in cui vengono formulate ipotesi, situato tra introduzione e metodo, e il paragrafo conclusioni che segue i risultati³¹.
- b) *Rassegne della letteratura* (*review articles* in inglese): si tratta di una raccolta dei più importanti contributi prodotti in un preciso periodo e riferiti a un preciso settore di ricerca. Non si tratta di una passiva presentazione o recensione dei testi selezionati, bensì lo scopo è individuare nuove prospettive di ricerca sulla base di errori e inconsistenze delle precedenti pubblicazioni. Una rassegna della letteratura si compone solitamente di tre parti: introduzione, corpo e conclusione. Tali parti potrebbero a loro volta presentare ulteriori suddivisioni.
- c) *Articoli teorici*: mirano ad analizzare e sviluppare una teoria o un insieme di teorie. La loro struttura può variare, ma in genere comprendono un'introduzione, un paragrafo sullo stato dell'arte, l'analisi o soluzione proposta e infine le conclusioni.
- d) *Articoli metodologici*: presentano nuove tecniche di analisi di dati o l'impiego di specifiche apparecchiature. Questa tipologia non segue una struttura fissa ed è rivolta principalmente a un pubblico di esperti.
- e) *Studi di caso*: dall'inglese *case studies*, si tratta di lavori in cui si analizzano singole entità, come individui, gruppi o comunità. Tale tipologia è impiegata principalmente in ambito medico e legale. L'articolo presenta inizialmente un'analisi della letteratura esistente sull'argomento e successivamente viene descritto il caso trattato. Alla descrizione segue

³¹ Il paragrafo 4.2.2 presenta un approfondimento sui singoli paragrafi sopramenzionati.

l'interpretazione teorica. Le parti costituenti del testo sono: introduzione, analisi generale, situazione osservata, intervento e conclusione.

- f) *Altri tipi di articoli*: tra le ulteriori tipologie di articoli. Tra queste si distingue l'articolo editoriale: un testo di poche pagine che introduce il numero della relativa rivista e che tratta temi specifici.

Sulla base dell'elenco sopra riportato, risulta evidente la notevole varietà e ampiezza delle tipologie di articoli scientifici. L'autore sceglie la tipologia più adatta in funzione dell'argomento trattato e della funzione comunicativa del testo. Gli articoli scientifici analizzati nel presente studio appartengono a diverse delle tipologie menzionate, a testimonianza della varietà di approcci e finalità che caratterizzano la scrittura scientifica.

4.2.2 Componenti dell'articolo scientifico

La struttura di un articolo scientifico si caratterizza per la presenza di elementi ricorrenti e comuni a più testi appartenenti a questa tipologia che svolgono una funzione specifica. Luccio (2014: 22-30) individua tali componenti e ne descrive le principali caratteristiche come segue:

- a) *il frontespizio*: deve rispettare specifiche regole e contenere due informazioni essenziali: il titolo e l'autore o gli autori. Il titolo rappresenta la primissima informazione fornita al lettore e deve trasmettere in modo chiaro e sintetico il nucleo del contenuto del testo.
- b) *l'abstract*: costituisce un riassunto autosufficiente del contenuto dell'articolo e ha la funzione di informare e orientare il lettore, fornendo una sintesi chiara degli obiettivi, del quadro teorico e/o metodologico e dei principali risultati. Può seguire il modello IMRaD (*introduction, methods, results and discussion*), che prevede la presentazione degli elementi fondamentali della ricerca, pur con variazioni in base alla tipologia dell'articolo (cfr. Luccio 2014: 25-26);
- c) *le parole chiave*: tale sezione compare sotto il titolo o sotto l'abstract. Le parole chiave, in inglese *keyword*, rappresentano concetti rilevanti, ricorrenti nel testo ed esplicano gli argomenti trattati;
- d) *l'introduzione*: in questa sezione l'autore fornisce al lettore il contesto in cui si colloca lo studio e presenta il problema, gli obiettivi e le ipotesi, illustrando la struttura dell'articolo;
- e) *il metodo e i materiali*: l'autore presenta il disegno di ricerca, i partecipanti coinvolti e le loro caratteristiche rilevanti, le procedure di raccolta dei dati, gli strumenti utilizzati e le modalità di analisi, con l'obiettivo di garantire la trasparenza e la riproducibilità della ricerca;
- f) *i risultati*: l'autore presenta in sequenza logica i risultati ottenuti dalla ricerca;

- g) *la discussione*: fornisce l'interpretazione dei risultati ottenuti dalla ricerca, alla luce di quanto esposto precedentemente;
- h) *le conclusioni*: questa sezione costituisce un epilogo dello studio condotto nel testo e riporta spunti di riflessione emersi nel corso della ricerca;
- i) *la bibliografia*: comprende l'insieme delle opere citate dall'autore nel suo contributo e fornisce al lettore i riferimenti necessari per rintracciare le fonti delle affermazioni presentate.

Oltre alle componenti sopramenzionate, è possibile individuare altri elementi che appaiono frequentemente nei testi scientifici. Si tratta di elementi che “illustrano gli articoli”, tra questi: tabelle, grafici, diagrammi, mappe, disegni e fotografie.

4.3 Il linguaggio scientifico

Thielmann (2017: 548) definisce la scienza “ein kollektiver Handlungszusammenhang, der neugiergeleitete Fragen hinsichtlich eines Phänomenbereichs einer Beantwortung zuführt”³². Sulla base di questa definizione, si può dedurre che la ricerca scientifica nasca dalla volontà di mettere in discussione un oggetto scientifico e di guardarlo da una prospettiva diversa rispetto alle conoscenze e alle interpretazioni già espresse e consolidate. Il meccanismo di domande e risposte, che ne deriva, governa la ricerca stessa e si fonda sull'uso del linguaggio. Considerando, pertanto, il ruolo essenziale del linguaggio nell'attività scientifica, Thielmann (2014: 4) descrive i modi in cui la scienza ne fa uso. In particolare, nel “fare scienza”, il linguaggio è utile per:

- l'elaborazione di domande scientifiche;
- la determinazione dei risultati;
- il processo di interrogazione e risposta scientifica;
- il confronto tra le nuove conoscenze e quelle esistenti, e la conseguente affermazione delle prime sulle seconde.

È così possibile analizzare il ruolo del linguaggio nella formulazione di quesiti scientifici e nell'elaborazione delle risposte, evidenziando come le strutture linguistiche contribuiscano alla costruzione e alla diffusione del sapere. Espressioni come *etwas untersuchen* ed *etwas ein Verfahren unterziehen* sono impiegate per esporre le procedure tipiche della ricerca scientifica. Analogamente, espressioni linguistiche come *die Meinungen gehen diesbezüglich auseinander* e *ist*

³² “[...] un contesto di azione collettivo che porta una risposta a domande, mosse dalla curiosità, riguardanti un determinato campo di fenomeni” (trad. mia)

allgemein bekannt sono impiegate per illustrare il confronto e la comparazione di conoscenze nuove e conoscenze già esistenti (Thielmann 2017: 549-550).

Quanto illustrato finora risulta essere coerente con le osservazioni di Ehlich (2000: 55), che considera il linguaggio scientifico non solo come parte integrante della comunicazione scientifica, ma anche come suo presupposto e conseguenza. In questa prospettiva, l'uso del linguaggio nella ricerca scientifica non si limita alla mera trasmissione di contenuti specialistici, bensì rappresenta uno strumento costitutivo dell'attività scientifica stessa. In continuità con quanto affermato da Thielmann (2014), Ehlich (1993, 1995, 2000) sposta l'attenzione dal semplice uso del linguaggio nel "fare scienza" alla struttura linguistica stessa della scienza, individuando tre dimensioni fondamentali che ne descrivono le funzioni e gli strumenti principali:

- funzione gnoseologica, che comprende il ruolo epistemologico del linguaggio nella costruzione, organizzazione, conservazione e trasmissione del sapere;
- linguaggio scientifico quotidiano (*alltägliche Wissenschaftssprache*), ossia l'insieme dei mezzi linguistici che sostengono e strutturano l'attività scientifica stessa (cfr. 4.3.3.2);
- funzione eristica e relative strutture, intese da Ehlich (1993) come le strategie linguistiche mediante le quali le conoscenze consolidate vengono messe in discussione e nuove acquisizioni vengono affermate e legittimate all'interno della comunità scientifica. Nel linguaggio scientifico tedesco la funzione eristica si manifesta attraverso strutture che evidenziano la divergenza all'interno della comunità scientifica (*die Meinungen gehen weit auseinander*), problematizzando lo stato della ricerca (*bisherige Untersuchungen sind nicht ausreichend, dieses Thema ist bislang unzureichend erforscht*) e giustificando la necessità di un contributo originale (*Das Ziel dieser Arbeit ist daher [...]*) (cfr. Thielmann 2017).

In sintesi, i linguaggi scientifici si configurano come varietà linguistiche che hanno sviluppato le proprie strutture per scopi scientifici attraverso un processo consapevole di elaborazione linguistica da parte di ricercatori. La consapevolezza di tali caratteristiche è cruciale non solo per descrivere la comunicazione scientifica, ma anche per comprendere le difficoltà che emergono nell'acquisizione di tale linguaggio in ambito didattico. Tradizionalmente, lo studio del linguaggio scientifico è stato ricondotto alla ricerca sul linguaggio specialistico, concentrandosi soprattutto sugli aspetti terminologici (Graefen 2001: 193). Questa impostazione ha contribuito a trascurare per lungo tempo gli aspetti grammaticali e lessicali che caratterizzano la costruzione del discorso scientifico e che, come sottolinea la stessa studiosa, costituiscono un problema rilevante nel linguaggio scientifico, poiché non rientrano nell'ambito classico della grammatica e del lessico, inteso come vocabolario costituito da singole parole. Molti tratti del linguaggio scientifico vengono interpretati

come fenomeni stilistici e, di conseguenza, gli errori commessi dagli studenti stranieri in questo ambito vengono spesso definiti come “errori espressivi” e per questo ritenuti di scarsa rilevanza (Graefen 2001: 193-195). In realtà, tali aspetti rivestono un ruolo centrale nella piena padronanza della lingua scientifica, in quanto riflettono le convenzioni comunicative e cognitive proprie della comunità accademica.

4.3.1 Linguaggio specialistico e linguaggio scientifico

Thielmann (2014: 3) definisce il linguaggio scientifico come “ein Idiom zur Vermittlung fachlicher Inhalte”³³. Tale definizione, che pone l’accento sulla funzione di trasmissione di contenuti specialistici, potrebbe indurre a ritenere erroneamente che il linguaggio scientifico si esaurisca nella sola componente terminologica e che, di conseguenza, il linguaggio specialistico (*Fachsprache*) e il linguaggio scientifico coincidano pienamente. Questa ipotesi ha influenzato a lungo lo studio del linguaggio scientifico; tuttavia, tra questi due concetti sussiste una distinzione sostanziale. Come osserva Thielmann (2017: 547), le varietà specialistiche (*fachliche Varietäten*) emergono per soddisfare esigenze comunicative strettamente legate alla divisione sociale del lavoro. Per esempio, chi lavora in un settore tecnico ha esigenze di denominazione che non si riscontrano nell’uso del linguaggio quotidiano. Il linguaggio specialistico costituisce quindi una forma espressiva che ha una portata limitata (Thielmann 2017: 547).

Come precedentemente evidenziato, inizialmente gli studi sul linguaggio scientifico si sono concentrati prevalentemente sui linguaggi specialistici. In Germania, questo filone di ricerca ha ricevuto un forte impulso grazie ai sostegni statali degli anni ‘70. Già Ammon (1973: 29) aveva sottolineato l’esistenza di tanti linguaggi specialistici quanti sono gli ambiti disciplinari. Successivamente, come osserva Thielmann (2017: 547), è stato Harald Weinrich (1990, 1995) a richiamare l’attenzione sulla necessità di considerare il linguaggio scientifico come un ambito autonomo. Le riflessioni di Weinrich si basano sulla constatazione che, sebbene la comunicazione scientifica, per via dei suoi contenuti, sia sempre di natura specialistica, essa non persegue finalità limitate a un singolo ambito disciplinare (Thielmann 2017: 547). In altre parole, mentre il linguaggio specialistico si concentra principalmente sulla denominazione e sull’uso di termini tecnici funzionali a un settore specifico, il linguaggio scientifico comprende un insieme molto più ampio di caratteristiche utili alla strutturazione e alla trasmissione del sapere in modo efficace, indipendentemente dai temi trattati. Per questo motivo, esso si configura come una varietà autonoma che, pur attingendo a termini e vocaboli settoriali, li rielabora all’interno di una

³³ “Un idioma per la trasmissione di contenuti specialistici” (trad. mia)

comunicazione orientata alla produzione e diffusione della conoscenza, come è stato illustrato sopra (cfr. 4.3) nella descrizione delle funzioni e degli impieghi del linguaggio scientifico.

4.3.2 Anglicizzazione vs plurilinguismo

“Purificare e arricchire la lingua madre è il compito delle menti migliori”, affermava Johann Wolfgang von Goethe (1875: 114). Attualmente questa citazione potrebbe non solo essere messa in discussione, ma addirittura risultare superata in quanto l’inglese si è imposto come lingua franca per eccellenza e la sua preponderanza è evidente in tutti i campi, compreso quello accademico e scientifico.

Come sostiene Reinbothe (2011: 18), la lingua inglese ha iniziato ad affermarsi come lingua dominante e, talvolta, come unica lingua della scienza nella comunicazione internazionale a partire dagli anni ‘60. Tuttavia, le cause che hanno portato a questa propensione per l’inglese rispetto ad altre lingue nazionali possono essere ricercate già nei decenni precedenti, in particolare nel periodo successivo alla Prima Guerra Mondiale. Gli assetti politici sviluppatisi conseguentemente al primo conflitto mondiale, il boicottaggio della lingua tedesca a livello scientifico internazionale e la successiva propaganda nazionalsocialista, di cui si faceva carico la lingua tedesca e la comunità tedesca scientifica, hanno contribuito alla marginalizzazione del tedesco e, parallelamente, al consolidamento dell’inglese come lingua della scienza (Reinbothe 2011: 9-18).

Oggi, questa supremazia linguistica si riflette anche nei sistemi formativi: per esempio numerose università, sia in Italia che in Germania, offrono corsi di laurea esclusivamente in inglese. Ne consegue una percezione della lingua scientifica tedesca come forma linguistica obsoleta, gradualmente sostituita dall’inglese, che viene considerato più moderno ed efficace nella comunicazione accademica. Tuttavia, tale fenomeno non riguarda soltanto il tedesco come linguaggio scientifico, ma anche le altre lingue nazionali, come il francese, l’italiano, lo spagnolo.

Secondo Ehlich (2000: 60-62), il passaggio da un multilinguismo scientifico a una condizione di monolinguisimo anglofono comporta conseguenze negative su più livelli:

- in primo luogo, comporta un impoverimento delle lingue delle singole comunità linguistiche. L’utilizzo della lingua nazionale nella comunicazione scientifica consente di coinvolgere la popolazione in questione nella comunicazione e di ampliare conseguentemente il raggio di comprensione e uso della lingua stessa;
- in secondo luogo, limita lo sviluppo della conoscenza del mondo, che trae beneficio dalle diverse prospettive offerte da ciascun linguaggio scientifico. Ehlich (2000: 54) evidenzia come ciascun linguaggio scientifico rappresenti sia il risultato che lo strumento di diverse forme di conoscenza scientifica e delle sue modalità di acquisizione. Pertanto, ogni

linguaggio scientifico non solo facilita la riflessione, ma ne costituisce anche la fonte. In quest'ottica, la traduzione e l'apprendimento di un linguaggio scientifico straniero assumono un ruolo essenziale per il progresso della scienza;

- infine, l'interazione tra diversi linguaggi scientifici rappresenta un'esperienza imprescindibile per la crescita intellettuale e la riflessione dei ricercatori. Nel contesto della ricerca scientifica internazionale il monolinguismo, al quale spesso si aspira, limita lo sviluppo della conoscenza, mentre l'uso di differenti linguaggi scientifici tutela la ricerca dall'etnocentrismo scientifico.

Sulla base di quanto esposto, si può dedurre che l'uso dell'inglese come lingua franca scoraggi le traduzioni, limitando la partecipazione ai dibattiti scientifici internazionali (Navarro et al. 2022: 4-5). La tendenza al monolinguismo anglofono comporta, dunque, implicazioni rilevanti anche sul piano dell'equità e dell'inclusione nella comunità scientifica globale. Come evidenziano Navarro et al. (2022: 3-5), l'uso esclusivo dell'inglese come lingua franca non garantisce automaticamente l'accesso paritario alla produzione e alla diffusione del sapere, ma tende piuttosto a rafforzare la centralità dei Paesi anglofoni post-industrializzati, a scapito delle aree considerate “periferiche”. In tal senso, uno studioso che si trova in un Paese anglofono post-industrializzato è potenzialmente in grado di condurre ricerche su un'ampia gamma di argomenti che riceveranno probabilmente maggiore attenzione rispetto agli studi di un accademico situato in un paese “periferico”, il quale, impiegando una lingua locale e una letteratura non inglese, riscontrerà maggiori difficoltà nel vedere la propria produzione riconosciuta come contributo legittimo al dibattito internazionale (Navarro et al. 2022: 4-5).

Inoltre, per i ricercatori non anglofoni, scrivere in inglese rappresenta spesso un onere aggiuntivo in termini di tempo, risorse e carico cognitivo, nonché un investimento economico in servizi di revisione e traduzione. Tale svantaggio strutturale può quindi limitare la loro produttività scientifica. Va anche considerato che il predominio dell'inglese può dare luogo a una distorsione nella valutazione del sapere: la conoscenza prodotta in lingua inglese tende a essere percepita come più autorevole rispetto a quella generata in altre lingue, che, anche se fiorente e in crescita, può essere ignorata, sottovalutata o esclusa più facilmente dalle principali discussioni scientifiche internazionali (Navarro et al. 2022: 3-4). Questa dinamica delinea una geografia della conoscenza fortemente asimmetrica, nella quale i centri di produzione e validazione della conoscenza coincidono con i Paesi anglofoni industrializzati, mentre gli altri vengono relegati a ruoli marginali. Il rischio che ne deriva per la scienza è la perdita della sua vocazione universale e pluralista.

In quest'ottica, la traduzione si configura come uno strumento imprescindibile per favorire la circolazione e l'accessibilità delle conoscenze scientifiche espresse in lingue diverse dall'inglese,

senza tuttavia sostituirsi ai linguaggi scientifici nazionali. Essa promuove l'emergere di risorse scientifiche linguistiche nazionali e allo stesso tempo è segno concreto dell'impegno e degli sforzi delle singole comunità linguistiche nell'affrontare le questioni chiave della scienza. Tali sforzi, seppur individuali, contribuiscono in maniera significativa allo sviluppo della comunità scientifica internazionale e all'ampliamento del patrimonio conoscitivo globale (Ehlich 1999: 11). In questa prospettiva, i diversi linguaggi scientifici nazionali costituiscono approcci complementari alla costruzione del sapere. Il compito della comunità scientifica internazionale dovrebbe essere quello di riconoscerne il valore, promuovendone l'uso e il continuo sviluppo, a tutela del pluralismo linguistico e cognitivo nella ricerca (Ehlich 1999: 11).

4.3.3 Le componenti del linguaggio scientifico

Come già delineato nei paragrafi precedenti, il linguaggio scientifico si connota per un'elevata densità strutturale, derivante dall'interazione di fenomeni lessicali e grammaticali specifici. Tale complessità può essere ricondotta, secondo Ehlich (1996/2007), a tre elementi principali, che rappresentano le dimensioni strutturali fondamentali:

1. elementi del linguaggio quotidiano (*Alltagssprache*);
2. elementi del linguaggio scientifico quotidiano (*alltägliche Wissenschaftssprache*);
3. elementi terminologici (*Fachterminologie*).

Diversamente dagli elementi del linguaggio quotidiano e della terminologia, che si possono incontrare anche al di fuori del linguaggio scientifico e in altri contesti comunicativi, il linguaggio scientifico quotidiano è un aspetto esclusivamente tipico del linguaggio scientifico. Di seguito tali dimensioni sono presentate in maniera più dettagliata.

4.3.3.1 Il linguaggio quotidiano e il linguaggio specialistico

Hoffmann (1988: 158) mette a confronto il linguaggio comune (*Gemeinsprache*, inteso come sinonimo di *Alltagssprache*) con il linguaggio specialistico e fornisce una definizione dei due termini: il linguaggio comune si riferisce alle parti del repertorio linguistico complessivo della società, impiegato in egual misura dai suoi membri in ambiti della vita simili per riferirsi a oggetti, fatti e idee generalmente noti. Il linguaggio specialistico, invece, include componenti linguistiche il cui uso negli ambiti della vita si differenzia tra i membri della società e sono impiegate per riferirsi a oggetti, fatti e idee specifiche dell'ambito trattato (Hoffmann 1988: 158).

Il confine tra linguaggio comune e linguaggio specialistico può sembrare e risultare estremamente sottile. Stolze (2013: 45) descrive il rapporto tra i due linguaggi come una relazione bipolare in cui si realizza uno scambio reciproco e dinamico. Tale interdipendenza deriva, da un

lato, dal fatto che il linguaggio specialistico si è sviluppato a partire dal linguaggio comune e, dall'altro, dall'assorbimento progressivo, da parte del linguaggio comune, di termini specialistici. Questo processo, che Stolze (2013: 49) definisce *Intellektualisierung der Gemeinsprache* (intellettualizzazione della lingua comune), riflette la percezione di modernità associata all'uso di termini specialistici. A livello lessicale, ciò si manifesta nella crescente presenza di termini specialistici, neologismi, parole composte e *Funktionsverbgefügen* (ad esempio *zum Ausdruck bringen*); a livello sintattico, si osserva una tendenza crescente alla sostantivizzazione (Stolze 2013: 47-49).

La Figura 8 illustra il rapporto tra il lessico del linguaggio comune (*Gemeinvokabular*) e il lessico del linguaggio specialistico (*Fachvokabular*).

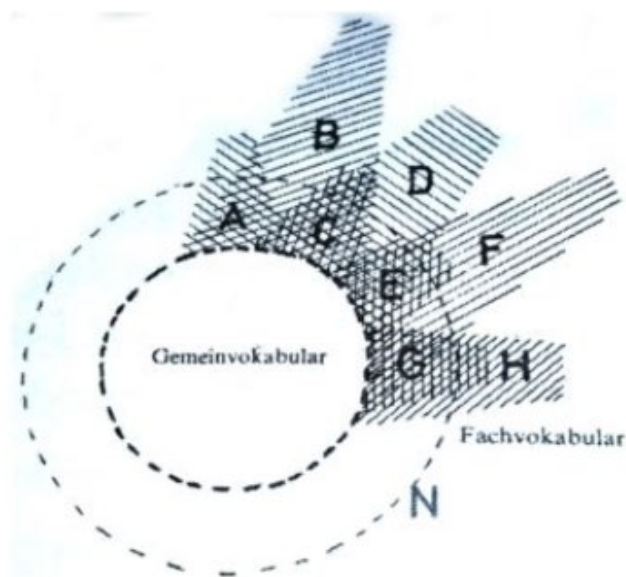


Figura 8: Lessico comune e lessico specialistico, Stolze (2013: 44)

Il cerchio centrale, etichettato come "Gemeinvokabular", rappresenta il lessico di uso generale, comprensibile alla maggior parte delle persone. Le sezioni esterne, indicate con "Fachvokabular" e segnalate da lettere (A, B, C, D, E, F, G, H), rappresentano vari ambiti del vocabolario specialistico. Le aree di sovrapposizione tra le diverse sezioni di "Fachvokabular" indicano i termini condivisi, mentre le sezioni dei "Fachvokabular" che non si sovrappongono rappresentano termini altamente specializzati, noti solo all'interno di un determinato dominio. Pertanto, il lessico comune costituisce la base sulla quale si sviluppa il lessico specialistico, più il lessico specialistico si allontana dal centro, più si specializza (Stolze 2013: 43).

Alla luce di quanto illustrato nella Figura 8, è possibile descrivere la differenza tra il linguaggio quotidiano e quello specialistico impiegando gli aggettivi "sfumato" e "netto". Paepcke (1980: 250) afferma che i linguaggi specialistici utilizzano termini che hanno confini linguistici

netti e marcati e ciò garantisce precisione e chiarezza nella trasmissione di conoscenze. Al contrario, il linguaggio comune utilizza termini che hanno un nucleo chiaro ma confini sfumati e non definiti, favorendo flessibilità comunicativa e comprensione tra parlanti con livelli di conoscenza diversi.

La compresenza di linguaggio quotidiano e linguaggio specialistico nei testi scientifici determina il grado di specializzazione linguistica (*Fachsprachlichkeitsgrad*). Kalverkämper (1990, cit. in Stolze 2013: 46) ha elaborato una scala, denominata *Skala der abnehmenden Fachsprachlichkeit*, che indica il grado di specializzazione linguistica dei testi, inteso come la quantità di caratteristiche linguistiche specialistiche presenti nel testo (cfr. Fig. 9).

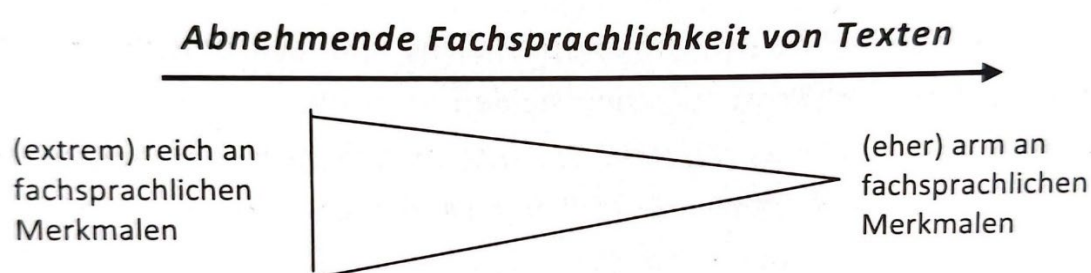


Figura 9: Grado di specializzazione linguistica, Stolze (2013: 46)

Alla sinistra della scala di Kalverkämper si collocano i testi estremamente ricchi di caratteristiche linguistiche specialistiche, destinati a una comunicazione strettamente specialistica tra studiosi ed esperti altamente specializzati. Questi testi sono caratterizzati da una comprensibilità limitata. All'estrema destra, invece, si trovano i testi di argomento tecnico, il cui linguaggio può essere classificato come comune. Tali testi sono poveri di caratteristiche linguistiche specialistiche e consentono quindi una comprensione più ampia (Stolze 2013: 46). Secondo Stolze (2013: 46), le caratteristiche del linguaggio specialistico che determinano il grado di specializzazione linguistica sono:

- terminologia tecnica specialistica (lessemi);
- caratteristiche linguistico-stilistiche (sintagmi, frasi, tipi di testo);
- elementi visivi come simboli, formule, grafici, illustrazioni.

Pertanto, il grado di specializzazione linguistica di un testo dipende dalla combinazione di lessico tecnico, struttura sintattica e supporti visivi, che insieme ne determinano la comprensibilità. A seconda di questo grado, si distinguono tre differenti tipi di comunicazione:

- alto grado di specializzazione linguistica,
- medio grado di specializzazione linguistica,

- basso grado di specializzazione linguistica.

Riprendendo la distinzione di Roelcke (2010: 57-58) basata su parole intraspecialistiche, interspecialistiche ed extraspecialistiche (*intrafachliche, interfachliche, extrafachliche Wörter*), è possibile individuare diversi tipi di comunicazione a seconda del grado di specializzazione. Un alto grado di specializzazione linguistica comporta una comunicazione intraspecialistica, in cui gli attori coinvolti sono degli esperti, si pensi ad esempio alla comunicazione tra due medici. Questo tipo di comunicazione è simmetrica e la comprensibilità è ridotta per i non esperti. Un medio grado di specializzazione linguistica connota una comunicazione interspecialistica, che coinvolge, ad esempio, un ingegnere e un operaio specializzato o due operai specializzati. Si tratta in questo caso di una comunicazione asimmetrica, poiché i partecipanti possiedono competenze diverse, pur risultando più comprensibile rispetto alla comunicazione intraspecialistica. Un basso grado di specializzazione linguistica definisce una comunicazione extraspecialistica, anch'essa asimmetrica, i cui attori possono avere un grado di conoscenza della materia trattata differente, come nel caso della comunicazione tra un medico e il paziente. In questo tipo di comunicazione, la comprensibilità è più ampia anche per i non esperti.

Questa differenziazione, basata sulle caratteristiche linguistiche dei testi, si ricollega a quanto esposto nel paragrafo 4.1 e alle differenziazioni che prevedono livelli di comunicazione scientifica differenti sulla base dei soggetti coinvolti.

4.3.3.2 Il linguaggio scientifico quotidiano

Il linguaggio scientifico quotidiano (*alltägliche Wissenschaftssprache*) è una componente della comunicazione scientifica, sia nella sua forma orale che scritta, che prevede l'uso specifico di parti del linguaggio quotidiano per gli scopi della scienza (Ehlich 2000: 52). La funzionalità di questi elementi è stata poco studiata poiché il linguaggio quotidiano rimane di frequente escluso dalla considerazione degli studiosi.

Graefen e Moll (2011: 17) sostengono che il termine “quotidiano” nella denominazione “linguaggio scientifico quotidiano” debba essere inteso in una duplice accezione. Si può trattare di:

1. parole impiegate non solo nel linguaggio scientifico, ma anche in quello quotidiano (ad esempio *die Diskussion, das Ergebnis, die Folge, der Unterschied*);
2. verbi e combinazioni idiomatiche, come le cosiddette *Fügungen*.

Feilke (1998) definisce quest'ultime come *idiomatische Fügungen*, cioè “costrutti idiomatici” composti da diversi lessemi, del linguaggio scientifico. La valenza semantica di tali combinazioni di lessemi non è desumibile dall'analisi delle singole componenti, poiché gli interi costrutti

costituiscono un'unità semantica autonoma (Graefen 2004: 293). Per via di questa peculiarità, tali combinazioni sono di frequente soggette a interpretazioni errate, da parte non solo di discenti stranieri, ma anche di parlanti nativi, i quali tendono a considerare i singoli lessemi separatamente e non nella combinazione data, senza individuarne, quindi, il significato complessivo e la funzione all'interno del testo.

Come indicato da Graefen (2004: 294), il principio che porta allo sviluppo di questi costrutti è l'uso. Tali combinazioni idiomatiche diventano "abitudini linguistiche" come conseguenza di un uso sempre più frequente da parte di una comunità di parlanti. Pertanto, in quanto parte del linguaggio scientifico quotidiano, le *Fügungen* danno forma ai testi scientifici tanto quanto il vocabolario specialistico.

L'uso consolidato di queste combinazioni di lessemi fa sì che essi ricorrano sempre insieme. Questa caratteristica riporta alla mente il concetto di *Kollokation*, collocazione. Tuttavia, come sottolinea Graefen (2004: 295), sussiste una differenza sostanziale tra *Kollokation* e *Fügung*. Il concetto di *Kollokation* si fonda su dati quantitativi di frequenza, ma tali criteri non risultano esaustivi per identificare il tipo specifico di costrutti denominati *Fügungen*. Una combinazione di lessemi diventa infatti una *Fügung* quando il parlante ne percepisce il valore semantico aggiunto, riconoscendola come unità semantica autonoma.

Per quanto riguarda la struttura delle *Fügungen* presenti nel linguaggio scientifico quotidiano, Graefen (2001, 2004) ne ha individuati due tipi:

- verbi e sostantivi che sono entrati a far parte di un vocabolario scientifico che si tramanda da anni, la cui origine va spesso ricercata in ambiti scientifici specifici. Spesso si tratta di termini specialistici generalizzati;
- lessemi del linguaggio comune che sono entrati a far parte del linguaggio scientifico grazie a un uso prolungato e frequente.

La prima categoria di lessemi, come indicato da Graefen (2001: 192), è costituita da termini che appartengono alla riflessione filosofica ed epistemologica. Un esempio di questo processo è il termine *Perspektive*, che risale alla scienza ottica del XVII e XVIII secolo (Graefen 1997). Una tipica *Fügung* con *Perspektive* è, ad esempio, *eine Perspektive auf X eröffnen*.

Per la seconda categoria di lessemi, l'origine va ricercata nel linguaggio quotidiano. La loro peculiarità consiste nel mutamento semantico che subiscono, pur mantenendo invariata la forma, quando si combinano con altri lessemi dando così origine a locuzioni fisse. Graefen (2001: 192) porta l'esempio della parola *Rahmen*, che acquista un significato specifico in ambito scientifico se combinata con il verbo *anwenden* (*einen Rahmen anwenden*).

Graefen e Moll (2011) hanno individuato differenti *idiomatische Fügungen* utilizzate nel tedesco scientifico scritto, ad esempio *Ziel der Arbeit ist (...)*, *Zusammenfassend lässt sich sagen, dass (...)*, *Die Erforschung konzentriert sich auf (...)*.

Nell'analisi del linguaggio scientifico quotidiano, è opportuno rilevare come non sussista un confine netto tra linguaggio comune e linguaggio scientifico quotidiano. Ciò deriva dall'impiego nel linguaggio scientifico quotidiano, di elementi propri del linguaggio quotidiano, che tuttavia a volte possono risultare inadeguati agli scopi scientifici. Graefen (2004: 297-298) esemplifica questa dinamica attraverso la distinzione tra i lessemi *Auffassung* e *Meinung*, spesso impiegati nel linguaggio comune come sinonimi. In ambito scientifico, tuttavia, emerge una differenza sostanziale: il primo appartiene al linguaggio scientifico quotidiano, il secondo no. La parola *Meinung* è strettamente associata alla dimensione della soggettività e della libertà di scelta, pertanto il suo utilizzo in un testo scientifico rischierebbe di inficiare la credibilità dei risultati presentati.

L'esempio analizzato dimostra come la scelta lessicale, anche all'interno del linguaggio scientifico quotidiano, richieda un controllo rigoroso affinché il testo mantenga chiarezza, coerenza e credibilità. La padronanza delle *Fügungen* e delle sfumature semantiche dei lessemi impiegati rappresenta, quindi, una competenza indispensabile per garantire l'efficacia e la credibilità della comunicazione scientifica.

4.3.4 Il tedesco come linguaggio scientifico

Lo sviluppo del tedesco come linguaggio scientifico è avvenuto con un relativo ritardo rispetto all'inglese, al francese o all'italiano, in parte a causa del ruolo del latino, quale lingua della scienza nel Medioevo e nell'epoca della scolastica. La stretta associazione del latino con la scolastica ha portato gli studiosi moderni di scienze naturali ad abbandonarlo nell'ambito della ricerca e degli studi. Se, infatti, la varietà scolastica del latino era funzionale ai filosofi naturalisti scolastici per la formulazione di argomenti logici, gli scienziati moderni, invece, necessitavano di un codice idoneo a veicolare non argomenti logici ma interpretazioni (Thielmann 2014: 6-7).

L'inizio dell'era moderna ha portato allo sviluppo di lingue scientifiche alternative al latino, come nel caso dell'inglese. Come linguaggio scientifico, quest'ultimo ha importato dal latino la componente terminologica, quindi tutto ciò che riguarda la funzione gnoseologica della lingua. Per quanto riguarda la componente eristica e le sue strutture, la varietà scolastica del latino risultava deficitaria e per questo motivo sono state impiegate le risorse della lingua comune inglese (Thielmann 2014: 7-10). Il processo di formazione del tedesco come linguaggio scientifico, invece, è stato estremamente differente: come evidenzia Thielmann (2014: 10), la lingua tedesca ha dovuto

prima elaborare la terminologia scientifica utilizzando le proprie risorse linguistiche, ritardandone quindi lo sviluppo.

Il vero padre della lingua tedesca della scienza è Christian Wolff, un filosofo illuminista, la cui missione intellettuale era rendere il sapere accessibile al pubblico che non conosceva il latino. Il suo lavoro di sviluppo della lingua ha preso avvio dalla creazione della componente terminologica; egli, tuttavia, si è dovuto anche confrontare con scienziati e filosofi che conoscevano esclusivamente le espressioni latine. Pertanto, per garantire l'accessibilità delle nuove conoscenze a questo specifico pubblico, ha creato elenchi di termini tedeschi con i corrispondenti termini latini (Thielmann 2014: 10). Ricken (1995: 61) cita alcune di queste neoformazioni tedesche:

- Anmerkung: scholion
- Aufgabe: problema
- Ausnahme: exceptio
- Aussage: enunciatio
- Begriff: notio, idea
- Eigenschaft: attributum
- Einbildungs-Kraft: imaginatio
- Erklärung der Sache: definitio realis
- Wort-Erklärung: definitio nominalis
- Grund-Satz: axioma

(Wolff 1726, cit. in Ricken 1995: 61)

Questi termini sono entrati a far parte della vita quotidiana, configurandosi come esempi del cosiddetto “linguaggio scientifico quotidiano”. Thielmann (2014: 11), invece, evidenzia come i principali processi di formazione delle parole in tedesco, tipici anche della lingua comune, abbiano avuto un ruolo fondamentale nel linguaggio scientifico. Tra i processi più rilevanti si annoverano:

- nomi composti, come *Einbildungskraft*, *Grund-Satz*;
- derivazioni deverbali, come *Aussage*, *Begriff*;

Questi due procedimenti, storicamente rilevanti, risultano tutt'oggi ancora produttivi sia nel linguaggio scientifico sia in quello comune, consentendo l'introduzione continua di nuovi termini per descrivere concetti emergenti.

Lo sviluppo del linguaggio scientifico tedesco è stato inoltre fortemente influenzato dal contesto storico e politico del Novecento, che ne ha determinato prima l'ascesa e poi il declino. All'inizio del XX secolo, la Germania ha beneficiato della creazione di numerosi istituti di ricerca

che hanno contribuito a una fiorente produzione scientifica in lingua tedesca e quindi alla diffusione internazionale della lingua tedesca come linguaggio scientifico (Reinbothe 2011: 1-2). Tuttavia, la Prima guerra mondiale ha rappresentato un punto di rottura. A causa del boicottaggio imposto dalle Accademie delle Scienze dei Paesi alleati, gli scienziati tedeschi e la lingua tedesca sono stati esclusi da congressi, associazioni e pubblicazioni internazionali. Il boicottaggio era giustificato dal sostegno degli studiosi tedeschi all'aggressiva politica bellica tedesca e dalla negazione dei crimini di guerra commessi dalle forze armate tedesche in Belgio e in Francia (Reinbothe 2011: 9). Il boicottaggio è stato formalmente interrotto nel 1926, tuttavia le trattative per la reintegrazione dei ricercatori tedeschi e della loro lingua in ambito scientifico si sono prolungate fino agli anni Trenta in seguito alle richieste di pari trattamento linguistico rispetto all'inglese e il francese. Ciò ha contribuito al declino dello status del tedesco come lingua per la comunicazione scientifica, accelerato anche dall'ascesa dell'inglese, favorita dalle fondazioni americane e dal ruolo centrale degli Stati Uniti nella ricerca. È interessante notare che, come risposta a questi fenomeni di marginalizzazione, in quegli stessi anni in Germania furono istituite diverse fondazioni e associazioni per la promozione della scienza e della lingua tedesca (Reinbothe 2011: 11-12).

Un altro punto di svolta è stato l'avvento del nazionalsocialismo e la Seconda Guerra Mondiale. La crescente politicizzazione della scienza e l'utilizzo ideologico della lingua tedesca in funzione propagandistica hanno provocato una reazione di rigetto da parte della comunità accademica internazionale. Come sostiene Reinbothe (2011: 16), molti studiosi si sono dissociati da istituzioni e riviste tedesche e hanno rifiutato l'uso della lingua tedesca per protestare contro la politica repressiva e antisemita del Nazionalsocialismo. Questo contesto sociopolitico, quindi, ha contribuito ulteriormente al declino dell'uso della lingua tedesca come linguaggio scientifico.

Nel secondo dopoguerra, pur in assenza di un boicottaggio formale come era accaduto nel primo conflitto mondiale, il tedesco non è riuscito a recuperare il proprio prestigio internazionale. Al contrario, ha perso definitivamente terreno a favore dell'inglese, che è diventata la lingua dominante, e quindi lingua franca, in convegni, pubblicazioni e nella comunità scientifica internazionale (Reinbothe 2011: 18-19).

Per quanto concerne le caratteristiche stilistiche e linguistiche del linguaggio scientifico tedesco, esse riflettono la necessità che il testo risulti il più possibile comprensibile, anche per un lettore non esperto della materia. Tale esigenza impone all'autore una presentazione e un'elaborazione accessibile e chiara dei contenuti complessi. Come in altri linguaggi scientifici il testo deve essere preciso e chiaro, semplice e conciso: la comprensibilità costituisce dunque un requisito fondamentale. La semplicità in un testo scientifico si manifesta attraverso frasi brevi, trattazioni non eccessivamente complesse, argomentazioni concrete e chiare, parole familiari e

spiegazioni di termini specialistici. La concisione, invece, si riferisce alla capacità del testo di trasmettere informazioni in proporzione alla sua lunghezza: un testo eccessivamente lungo può risultare difficile da comprendere, poiché si potrebbe perdere l'attenzione durante la lettura. Tuttavia, anche in un testo troppo breve potrebbero verificarsi problemi di comprensione, ad esempio a causa dell'omissione di informazioni indispensabili. Pertanto, per ottenere un testo conciso, si devono distinguere gli elementi essenziali dagli elementi non essenziali, è necessario omettere i dettagli non significativi, eliminare gli aggettivi superflui e le parole riempitive. Bisogna concentrarsi sulla domanda di ricerca e limitarsi all'essenziale. Anche il rispetto delle regole di base della grammatica e dell'ortografia rientrano nelle caratteristiche di uno stile comprensibile.

Un ulteriore requisito fondamentale del testo scientifico è l'obiettività. Come evidenziano Graefen e Moll (2011: 99-120) riferendosi al testo scientifico tedesco, l'autore adotta precise strategie stilistiche per soddisfare tale esigenza. Queste sono:

1. *“Ich-Verbot”* (divieto dell'uso della prima persona): l'impiego della deissi personale (*io, mio, noi, nostro*) nel linguaggio scientifico rappresenta una questione ampiamente dibattuta. L'autore tende a porsi in secondo piano al fine di garantire un'esposizione il più possibile oggettiva; di conseguenza l'impiego della deissi personale, in particolare della prima persona singolare “io” viene solitamente evitato nella scrittura scientifica. Tuttavia, in due casi il ricorso a tale alla deissi personale si rivela funzionale: 1) se l'autore deve riferire di un'analisi o di un esperimento che ha condotto con altri studiosi, può usare il pronome “noi” riferendosi al gruppo di ricercatori; 2) se l'autore vuole includere se stesso e i lettori in un “gruppo di parlanti” può impiegare il *noi inclusivo*, ad esempio *Wir werden im Folgenden die wichtigsten Ergebnisse zusammenfassen*).
2. Uso della diatesi passiva al posto dell'attiva: nel linguaggio scientifico le costruzioni passive sono impiegate con maggior frequenza rispetto a quelle attive, poiché consentono un'esposizione più impersonale e generale dei contenuti. Nel passaggio dalla forma attiva a quella passiva, l'agente dell'azione viene messo in secondo piano, contribuendo così a rendere l'enunciato più oggettivo³⁴. Un ulteriore vantaggio dell'impiego del passivo risiede nella possibilità di enfatizzare maggiormente un oggetto, un fatto o un'azione, poiché questi diventano il soggetto grammaticale, come si nota dal seguente esempio che mostra il passaggio dalla forma attiva a quella passiva: *Der Forscher untersucht die sprachlichen Merkmale des Textes.* > *Die sprachlichen Merkmale des Textes werden untersucht.*

³⁴ Si noti che la frase passiva non è ammessa per i verbi riflessivi.

3. formulazioni impersonali: l'ampio ricorso a questi costrutti risponde all'esigenza di conferisce al testo scientifico un elevato grado di generalità (ad esempio *es bedarf; es ist anzunehmen, dass*).
4. stile nominale: nei testi scientifici si prediligono nomi e aggettivi rispetto ai verbi. I sostantivi diventano quindi i principali vettori del carico informativo, permettendo una notevole densità semantica (ad esempio *Aufgrund des Fehlens der Daten ist eine genaue Analyse nicht möglich.*).

Oltre a quanto menzionato precedentemente, nello stile scientifico rientrano altre precise scelte stilistiche, ad esempio l'autore deve evitare ripetizioni ed espressioni personali. È inoltre essenziale mantenere la coerenza terminologica, evitando di utilizzare termini diversi per indicare uno stesso oggetto. Per quanto concerne i tempi verbali, si raccomanda di non effettuare frequenti passaggi da un tempo verbale all'altro; inoltre è preferibile l'uso del tempo perfetto e del presente.

Nella loro pubblicazione Graefen e Moll (2011: 119-120) trattano anche il corretto uso del *Konjunktiv II*. Nei testi scientifici tedeschi, esso deve essere impiegato esclusivamente per esprimere una supposizione o per indicare l'oggetto di cui si parla come probabile causa o conseguenza ipotetica. È da ritenersi errato l'impiego del *Konjunktiv II* per introdurre un esempio o per marcare una considerazione e/o riflessione come ipotesi o raccomandazione. In entrambi i casi, l'uso del tempo presente è giudicato più corretto.

In sintesi, il presente paragrafo ha illustrato le caratteristiche del linguaggio scientifico tedesco, tuttavia molte delle peculiarità discusse si riscontrano anche in altri linguaggi scientifici, come sarà approfondito nel paragrafo successivo dedicato all'italiano.

4.3.5 L'italiano come linguaggio scientifico

Come già messo in luce nel paragrafo 4.3.4, l'evoluzione dell'italiano come linguaggio scientifico precede lo sviluppo del linguaggio scientifico tedesco. Le sue origini possono essere ricercate più nello specifico nel 1600 (cfr. Bianchi 2020; Migliorini 2001).

In Italia il latino ha assunto un ruolo predominante nell'ambito scientifico e lo ha mantenuto per secoli: l'insegnamento universitario avveniva esclusivamente in latino e le trattazioni scientifiche e filosofiche erano redatte in questa stessa lingua. Tuttavia, durante il XVII secolo, emerge una nuova tendenza, iniziata da Galileo Galilei. Nel 1611 lo studioso scrive la prima lettera in italiano volgare indirizzata a Mons. Dini. Successivamente, compone altre lettere e le sue opere maggiori in italiano (Migliorini 2001: 392). Dietro la scelta di Galileo, si cela una specifica motivazione ideologica e polemica: lo studioso intende staccarsi dalla vecchia scuola, chiusa e senza contatti con la realtà, per avvicinarsi a un pubblico "nuovo".

Nei *Discorsi e dimostrazioni intorno a due nuove scienze* (1638), Galileo sostiene che le verità naturali devono essere comprese dal maggior numero possibile di persone e per raggiungere tale scopo la lingua impiegata non deve essere troppo lontana da quella comune. Questa sua scelta lo espone a numerose critiche: prima fra tutte quella di Keplero che lo accusa di *crimen laesae humanitatis*, un crimine contro l'umanità, inteso come crimine contro la scuola classica. Molte critiche provengono anche da editori stranieri che continuano a richiedere le traduzioni latine delle sue opere (Migliorini 2001: 393). Nelle sue pubblicazioni Galileo impiega termini della lingua italiana comune, non greca o latina come i suoi colleghi studiosi, sottraendoli al loro uso quotidiano e attribuendo loro significati scientifici precisi. Un esempio è rappresentato dalla parola "pendolo" che in fisica indica un oggetto solido che, appeso a un filo o a un'asta, oscilla se mosso dall'alto. Galileo ricorre alla nominalizzazione di un aggettivo già presente nella lingua comune per indicare un oggetto sospeso e in movimento. Tra le parole della fisica e dell'astronomia coniate da Galileo o da lui adoperate in accezione specifica, si ricordano ad esempio forza, velocità, momento e resistenza (Dardano 1994: 533-534). In seguito all'operato di Galileo e dei suoi diretti discepoli, l'italiano inizia ad affiancare il latino nelle pubblicazioni sulle scienze naturali e mediche, nella legislatura e nella vita pubblica, ad esempio nelle recite teatrali (Migliorini 2001: 393-394).

Per quanto concerne gli aspetti linguistici e stilistici del linguaggio scientifico italiano, è possibile ribadire quanto già osservato in relazione al linguaggio scientifico tedesco nel paragrafo 4.3.4. L'autore, all'atto della redazione di un testo scientifico, deve rispettare dei requisiti fondamentali: il testo deve essere preciso, senza ambiguità e conciso. Si devono esporre i contenuti con chiarezza, impiegando il numero minore di parole possibili. Il testo deve essere privo di errori grammaticali, sintattici e lessicali.

Cortelazzo (1994: 5-6) individua i seguenti tratti caratteristici del linguaggio scientifico italiano a livello morfosintattico:

- riduzione di tempi, modi, persone verbali a favore della terza persona dell'indicativo presente;
- ricorrente uso di verbi generici (ad esempio essere, avvenire, riferirsi). L'uso di questi verbi permette di dare risalto ai sostantivi e ai sintagmi nominali che li accompagnano;
- ricorso frequente a forme nominali del verbo: in particolare i participi passato e presente tendono a perdere la loro funzione verbale originaria per assumere valore nominale o aggettivale (ad esempio "Si consideri il soggetto X avente...");
- prevalenza dello stile nominale: i sintagmi verbali sono sostituiti da nominalizzazioni e sintagmi nominali;
- prevalenza di forme passive e forme impersonali.

L'ultimo tratto menzionato permette in particolar modo il rispetto di un ulteriore requisito essenziale della scrittura scientifica: l'oggettività. Attraverso l'impiego di forme impersonali e della diatesi passiva, l'autore tende a rimanere sullo sfondo, lasciando emergere in primo piano l'azione e/o l'oggetto dell'enunciato.

Nel complesso, il testo scientifico si caratterizza per requisiti stilistici e linguistici che prescindono dalla specifica lingua scientifica utilizzata e che risultano pertanto comuni a più contesti linguistici. Ciò che invece varia in base alla lingua di riferimento sono le caratteristiche morfosintattiche. Il quadro generale tracciato nel presente capitolo costituisce il fondamento teorico per l'analisi sviluppata nel capitolo successivo, dedicato allo studio empirico, fulcro del presente lavoro, che si concentra in particolare sulle sfide traduttive e sul ruolo del *post-editing* nella resa in italiano del linguaggio scientifico tedesco, con particolare riferimento agli articoli scientifici.

Studio empirico: traduzione con DeepL su Trados Studio

Il presente capitolo si concentra sulla descrizione dello studio empirico volto ad analizzare la resa in lingua italiana di articoli scientifici in tedesco inerenti alla linguistica tedesca e tradotti con il sistema di TA DeepL. A tal fine sono stati costituiti due corpora paralleli: uno in lingua tedesca e l'altro in lingua italiana che vengono presentati in apertura. Successivamente vengono illustrati gli strumenti utilizzati: DeepL, Trados Studio e Sketch Engine. Il capitolo prosegue descrivendo in dettaglio le fasi dello studio empirico, evidenziandone procedure e scelte operative. Si intende così documentare il processo sperimentale, offrendo una visione chiara e strutturata dell'iter metodologico seguito, a garanzia della replicabilità dello studio.

5.1 Materiali

I materiali sono costituiti da due corpora: un corpus di articoli scientifici in lingua tedesca e uno di articoli scientifici in lingua italiana, in entrambi i casi si tratta di testi originali. I due corpora sono comparabili: i testi sono stati selezionati in modo da appartenere allo stesso ambito disciplinare, quello della linguistica tedesca, e al medesimo genere testuale, l'articolo scientifico.

Il corpus tedesco costituisce la base per la generazione dei testi tradotti con DeepL ed è oggetto di valutazione dello studio. Si compone di articoli scientifici in lingua tedesca scelti da tre riviste scientifiche *open access* e riconosciute dall'ANVUR come riviste scientifiche di Classe A, *Informationen Deutsch als Fremdsprache*, *trans-kom* e *Interculture journal*, che trattano rispettivamente il tedesco come lingua straniera, la traduzione e la comunicazione interculturale. I contributi sono stati selezionati dopo attenta lettura e analisi finalizzata a individuare gli elementi chiave del linguaggio scientifico illustrati da Ehlich (1996/2007, cfr. 4.3.3) e presi in esame in questo progetto di ricerca, e cioè la terminologia, il linguaggio comune e il linguaggio scientifico quotidiano (*alltägliche Wissenschaftssprache*).

Prima di procedere alla fase traduttiva, sono stati rimossi dai contributi i riferimenti bibliografici, gli abstract in lingua inglese e le note biografiche in quanto non oggetto di studio. Si è deciso, invece, di mantenere eventuali grafici e tabelle per avere una panoramica più completa della traduzione automatica di contributi scientifici, oltre che per garantire maggiore chiarezza circa il contenuto del testo tradotto.

Si è deciso di selezionare contributi scientifici provenienti da tre fonti differenti, al fine di evitare una focalizzazione esclusiva su un singolo sottoambito specialistico e di ottenere,

conseguentemente, un quadro più ampio e rappresentativo della qualità della traduzione automatica nel campo degli studi linguistici. Le edizioni da cui sono stati selezionati i contributi appartengono a un arco temporale che va dal 2013 al 2022. Questa scelta è legata alla facilità di reperimento di materiale pertinente ai fini della ricerca. Il numero degli articoli ammonta a 27, precisamente 9 articoli per ogni fonte selezionata, per un totale di 221.167 *tokens*. Il totale dei *tokens* è distribuito come segue tra le tre riviste scientifiche: *Informationen Deutsch als Fremdsprache* 64.224 *tokens*; *trans-kom* 80.711 *tokens*; *Interculture journal* 76.232 *tokens*.

Nel caso del corpus italiano, questo funge da termine di confronto per l'analisi linguistica e stilistica degli output della TA nella fase di *post-editing*, permettendo l'individuazione di espressioni e collocazioni tipiche del linguaggio scientifico italiano. È composto da articoli scientifici in lingua italiana selezionati da diverse riviste scientifiche *open access* e riconosciute dall'ANVUR come riviste scientifiche di Classe A, scelte sulla base degli ambiti disciplinari trattati nel corpus di lingua tedesca, al fine di garantire la massima corrispondenza tematica tra i due insiemi di dati. In questo caso, i temi affrontati sono la didattica delle lingue straniera, la traduzione e l'interculturalità. Le riviste selezionate sono: *Educazione interculturale*, *EL.LE*, *inTRAlinea*, *Italiano a scuola*, *Italiano a stranieri – Edilingua*, *Italiano LinguaDue*, *Lingua culture mediazioni*, *Lingue e linguaggi*, *LTO – Lingue e testi di oggi*, *Mediazioni*, *Rivista internazionale di tecnica della traduzione*.

Le edizioni da cui provengono i contributi coprono un arco temporale che va dal 2013 al 2023. Per questo corpus si è deciso di adottare lo stesso intervallo temporale individuato per il corpus di lingua tedesca, includendo tuttavia anche il 2023 per ampliare l'estensione del corpus. Il numero degli articoli selezionati ammonta a 200, per un totale di 1.188.417 *tokens*. Una dimensione così ampia è stata ritenuta necessaria in quanto il corpus è destinato a supportare la fase di *post-editing*, per la quale è opportuno disporre di una base testuale sufficientemente ampia per garantire risultati affidabili e rappresentativi. Dagli articoli di questo corpus si è deciso di rimuovere riferimenti bibliografici, abstract in lingua inglese, note biografiche, figure, tabelle, grafici e didascalie che li accompagnano che non risultano rilevanti per la finalità di ricerca e quindi non sono funzionali ai fini dell'estrazione di elementi linguistici del linguaggio scientifico. Alla luce delle finalità dei corpora e per una gestione più agevole dei dati, si è scelto di impostarli sul software Sketch-Engine, strumento che sarà presentato successivamente nel presente capitolo (cfr. 5.2.3).

5.2 Strumenti

Per la realizzazione dello studio empirico sono stati impiegati tre strumenti principali, selezionati in base alle loro specifiche potenzialità rispetto agli obiettivi della ricerca: DeepL, utilizzato per la traduzione automatica del corpus in lingua tedesca; Trados Studio, impiegato per la gestione dei progetti di traduzione e per le operazioni di *post-editing*; Sketch Engine, adottato per l'analisi linguistica e la gestione dei corpora. Nei paragrafi successivi, ciascuno di questi strumenti verrà descritto in maniera dettagliata, soffermandosi sulle rispettive caratteristiche tecniche, sulle modalità di impiego nell'ambito della ricerca e sul ruolo svolto nelle diverse fasi del lavoro.

5.2.1 DeepL Translator

Secondo l'*ALC Industry Survey 2024*, condotto dall'Association of Language Companies (ALC) in collaborazione con Slator, DeepL rappresenta il principale fornitore di traduzione automatica a livello globale. L'indagine, realizzata tra il 17 giugno e il 29 luglio 2024 su un campione di 127 agenzie di servizi linguistici provenienti da 28 Paesi, evidenzia che l'82% delle agenzie utilizza DeepL, a fronte del 46% che impiega Google, del 32% Microsoft e del 17% Amazon AWS. Alla luce di questi dati e della rilevanza che questo strumento riveste sul mercato internazionale della traduzione, DeepL è stato scelto come oggetto di analisi. Non sono invece stati considerati ChatGPT e strumenti analoghi di intelligenza artificiale, poiché l'obiettivo della ricerca è esaminare la traduzione automatica prodotta da sistemi di TA e non da chatbot, che non operano come traduttori automatici bensì come generatori di testi.

5.2.1.1 Caratteristiche e componenti

DeepL Translator è un sistema di TA neurale lanciato nell'agosto nel 2017 dalla società tedesca DeepL GmbH con sede a Colonia. Il nome DeepL sta per *deep Learning* (DL): questo sistema di TA si basa infatti sul *deep Learning* e sulla TA neurale³⁵. DeepL costituisce un'evoluzione di Linguee, un dizionario online bilingue con esempi nella lingua di partenza e di arrivo, primo prodotto della società, inizialmente chiamata Linguee GmbH, e successivamente rinominata DeepL GmbH al momento del lancio di DeepL Translator.

Alla sua nascita, la tecnologia di DeepL sfruttava le reti neurali e si avvaleva della potenza di un supercomputer in Islanda, con una capacità di 5,1 petaFLOPS che poteva tradurre un milione di parole in meno di un secondo. Il supercomputer in questione veniva impiegato per addestrare le reti neurali servendosi di una vasta raccolta di testi multilingui. Durante l'addestramento, le reti esaminano le traduzioni e imparano a tradurre autonomamente le frasi avvalendosi della corretta

³⁵ La TA neurale è stata trattata dettagliatamente nel paragrafo 1.3.1.2 e nei rispettivi sottoparagrafi.

grammatica e delle corrette scelte lessicali (Schwan 2017). Nel marzo del 2025, DeepL ha annunciato una collaborazione con Nvidia, che gli ha consentito di dotarsi del supercomputer Nvidia DGX SuperPOD con sistemi DGX GB200³⁶. Grazie a questa nuova infrastruttura, l'azienda ha dichiarato di poter tradurre l'intero *Oxford English Dictionary* in soli 3,72 secondi, rispetto ai 39,06 secondi richiesti dalla versione precedente del sistema, e di essere in grado tradurre l'intero contenuto di Internet in 18,45 giorni, rispetto ai 193,76 giorni necessari con la tecnologia precedente. Questi tempi si ridurrebbero ulteriormente con un'aggiuntiva ottimizzazione del sistema³⁷.

Attualmente DeepL supporta 36 lingue, includendo le varianti dell'inglese americano e britannico, il cinese semplificato e tradizionale, lo spagnolo latino-americano e il portoghese brasiliano. La versione gratuita consente di tradurre fino a 50.000 caratteri al mese, creare un glossario composto da 5 voci, salvare una traduzione per continuare a lavorarci in un secondo momento e tradurre un file non modificabile al mese³⁸.

Da marzo 2018 è disponibile il servizio di abbonamento a pagamento DeepL Pro, che offre diverse funzioni aggiuntive rispetto alla versione gratuita, tra le quali la più rilevante è la garanzia di sicurezza dei dati. Questo significa che durante l'utilizzo da parte dell'utente, il sistema di TA non salva le informazioni che vengono immesse o create, impedendone, quindi, l'impiego per l'addestramento delle reti neurali. La sicurezza dei dati non viene garantita durante l'uso della versione gratuita, motivo per il quale si sconsiglia l'uso di quest'ultima versione per scopi lavorativi e/o per traduzioni che contengono dati sensibili.

La funzione di glossario è stata integrata a partire da agosto 2021 e consente di creare glossari manualmente o di importarli da file esterni. Questa funzione è attualmente disponibile per 30 lingue³⁹. Da gennaio 2023, inoltre, gli utenti possono usufruire di una nuova funzionalità: DeepL Write. Si tratta di un assistente di scrittura basato sull'IA, inizialmente disponibile solo per l'inglese e il tedesco. Allo stato attuale DeepL Write supporta anche il francese, l'italiano, il portoghese e lo spagnolo⁴⁰. Esiste inoltre una versione a pagamento di DeepL Write, DeepL Write Pro, che offre

³⁶ Per maggiori informazioni, si veda <https://www.nvidia.com/it-it/data-center/dgx-gb200/> (ultima consultazione 20/12/25).

³⁷ Ulteriori informazioni sono disponibili al seguente link: <https://www.deepl.com/en/blog/tech/how-to-deploy-nvidia-dgx-superpod> (ultima consultazione: 20/12/25).

³⁸ Dati aggiornati al 22/10/2025. Le caratteristiche del servizio e il numero di lingue supportate possono essere soggetti a modifiche successive.

³⁹ Dati aggiornati al 22/10/2025.

⁴⁰ Dati aggiornati al 22/10/2025. Per maggior informazioni, si veda <https://www.deepl.com/it/products/write> (ultima consultazione 22/10/25).

vantaggi quali la sicurezza dei dati, suggerimenti illimitati per il miglioramento dei testi, uso senza limiti di alternative e di stili di scrittura⁴¹.

Attualmente DeepL Pro presenta tre differenti piani a pagamento: *Individual*, *Team* e *Business*. Il piano *Individual*, oltre a garantire la sicurezza dei dati come tutti i piani di DeepL Pro, offre un limite massimo di 300.000 caratteri al mese, consente di tradurre tre testi di file modificabili e pagine intere di testi. Questo piano consente l'utilizzo del tool a un solo utente. Per il piano *Team*, il limite mensile di caratteri è fissato a 1.000.000. È possibile tradurre fino a 20 testi di file modificabili al mese, si possono eseguire traduzioni di file in blocco, salvare 10 traduzioni e creare fino a 5 glossari da 100 voci. Il piano *Team*, così come quello *Business*, consente l'uso a un numero illimitato di utenti e permette inoltre l'integrazione del sistema di TA in diversi CAT tools, tra i quali Trados Studio e memoQ. Il piano *Business* si differenzia da quello *Team* perché non presenta un limite di caratteri da tradurre mensilmente, è possibile tradurre fino a 100 file modificabili al mese, salvare un numero illimitato di traduzioni e creare fino a 250 glossari con voci senza limiti. In aggiunta, il piano *Business* include l'utilizzo di DeepL Write Pro. Dal punto di vista dei costi, i piani differiscono a seconda delle funzionalità e del numero di utenti: *Individual* è il più economico ed è adatto per chi lavora da solo, mentre *Team* e *Business*, essendo destinati a più utenti, comportano costi totali superiori, ma offrono anche maggiori funzionalità, come già illustrato sopra⁴². Non essendo inclusa la funzionalità DeepL Write Pro nei piani *Individual* e *Team*, è possibile usufruire di due abbonamenti appositi che combinano i vantaggi dei rispettivi piani con le funzionalità dello strumento di scrittura assistita⁴³.

I tre piani a pagamento sopra descritti sono il risultato di un recente aggiornamento. Quando DeepL è stato integrato in questo studio di ricerca, presentava tre piani a pagamento differenti: *Starter*, *Advanced* e *Ultimate*. La distinzione più rilevante tra i piani precedenti e quelli attuali riguarda il limite mensile dei caratteri traducibili: nei piani a pagamento precedenti non era previsto alcun limite, consentendo la traduzione di un numero illimitato di caratteri. Il numero di file modificabili traducibili al mese rimane invariato, così come la possibilità di integrare il sistema di TA nei CAT tools. Le differenze riguardano il numero di glossari: il piano *Starter* ne permetteva solo uno, mentre i piani *Advanced* e *Ultimate* fino a 2.000.

5.2.1.2 Come funziona DeepL?

L'accesso a DeepL può avvenire tramite il sito web o scaricando l'applicazione desktop, disponibile sia per Windows che per macOS. È inoltre disponibile un'applicazione per dispositivi mobili.

⁴¹ Per maggiori informazioni, si veda <https://www.deepl.com/it/pro> (ultima consultazione 22/10/25).

⁴² Cfr. nota 36.

⁴³ Cfr. nota 36.

Di seguito viene spiegato il funzionamento del servizio mediante sito web, accessibile all'indirizzo <https://www.deepl.com/translator>. L'interfaccia presenta due campi principali. Nel riquadro a sinistra viene inserito il testo da tradurre, sia digitando tramite tastiera sia utilizzando funzione copia e incolla. È possibile caricare anche un documento in quanto il sistema di TA supporta diversi formati, tra i quali .pdf, .docx e .pptx. Nel riquadro a destra invece appare la traduzione generata da DeepL. La lingua di partenza viene riconosciuta automaticamente, o può anche essere selezionata nell'elenco di lingue offerte dal sistema di TA, analogo a quello mostrato nella Fig. 10 per la lingua di arrivo.

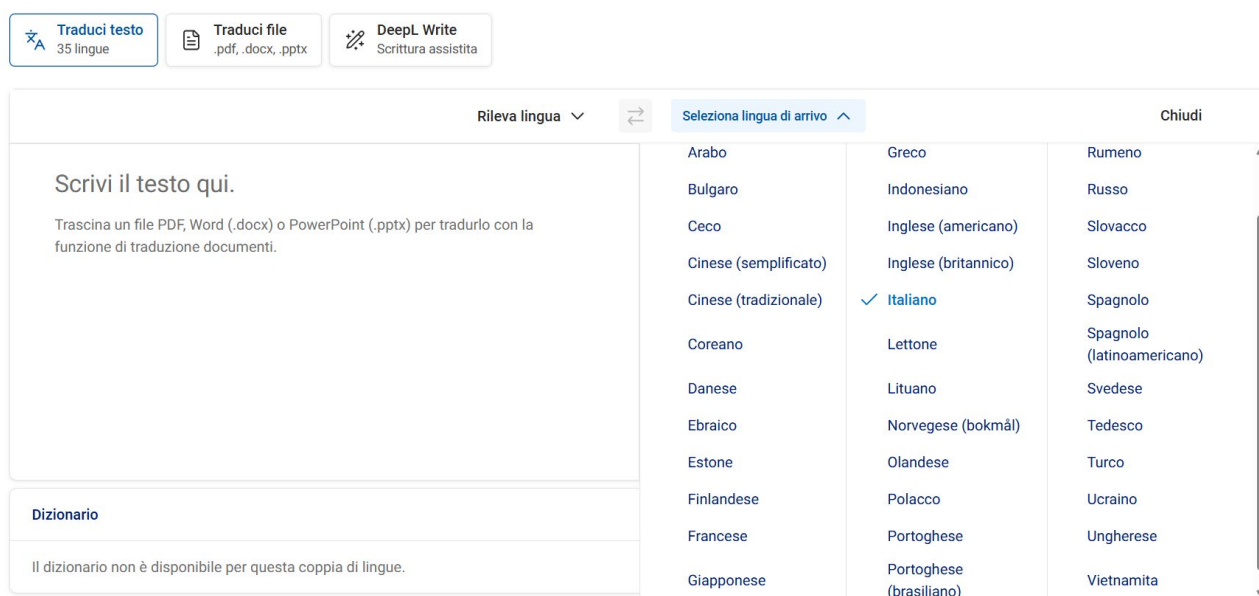


Figura 10: Interfaccia di DeepL n.1

Nel riquadro a destra, nell'angolo in alto destra, si trova il tasto "Opzioni". Cliccando, l'utente ha accesso a due funzionalità: la prima consente di scegliere lo stile della traduzione, formale o informale; la seconda riguarda la funzione glossario. Queste due ultime funzioni sono disponibili esclusivamente nella versione a pagamento, DeepL Pro, come già menzionato precedentemente. L'interfaccia utente consente inoltre il *post-editing*: per ogni parola tradotta da DeepL è possibile selezionare una traduzione alternativa, scegliendo tra parole o espressioni diverse (cfr. Fig. 11).

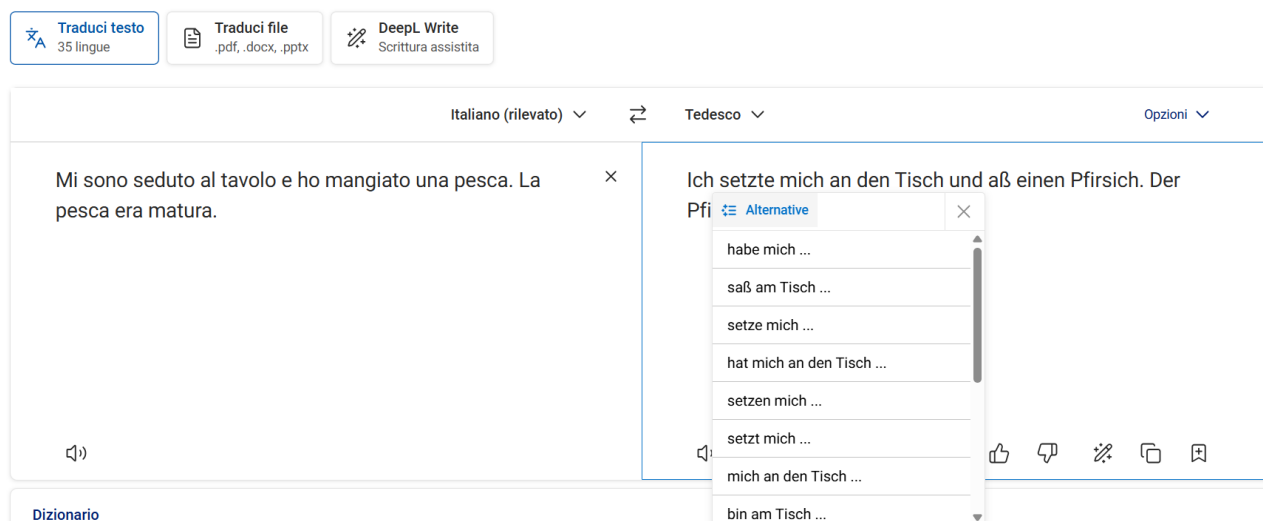


Figura 11: Interfaccia di DeepL n.2

Per il presente progetto di ricerca si è deciso di impiegare il piano tariffario DeepL Pro Advanced. Questa scelta è motivata dal fatto che, a differenza del piano gratuito e degli altri piani tariffari a pagamento, non impone un limite massimo di caratteri nella traduzione, offre maggiore sicurezza dei dati e permette l'integrazione in numerosi CAT tools, come si è deciso di operare in questo studio.

5.2.2 Trados Studio

Trados Studio è lo strumento di traduzione assistita che copre gran parte del mercato: secondo Lecci e Di Bello (2012: 12-13) è utilizzato “da più dell’80% dei traduttori professionisti in tutto il mondo”. Per il presente progetto di ricerca si è scelto di impiegare tale tool, poiché consente l'integrazione del sistema di TA selezionato, combinando così i vantaggi della TA neurale con le potenzialità caratteristiche dei CAT tool. Inoltre, Trados Studio mette a disposizione degli utenti due plug-in rilevanti per la fase di *post-editing*. Il primo è “Post-Edit Compare” che permette di confrontare due versioni della traduzione, quella originale generata dalla TA e quella post-editata, e segnala le modifiche apportate nella fase di *post-editing*. Il secondo plug-in è “Qualitivity” che permette di tracciare il tempo impiegato per svolgere la traduzione e il *post-editing* dei segmenti contenuti nel documento, oltre a tracciare ogni singola modifica apportata ai segmenti a livello granulare. Entrambi i plug-in citati risultano quindi particolarmente funzionali per la valutazione del *post-editing effort*. Integrando DeepL Pro a Trados Studio, il sistema di TA effettuerà la traduzione direttamente sul CAT tool e questa sarà visualizzata nell'editor di testo, dove potrà essere modificata nella fase di *post-editing*. La versione impiegata nel presente studio è Trados Studio 2022 SR2.

5.2.2.1 Componenti essenziali di Trados Studio

Le componenti essenziali che caratterizzano tutti i CAT tool, e quindi anche Trados Studio, memoria di traduzione, database terminologico e editor di testo, sono state presentate già nel paragrafo 1.2.1.

Le memorie di traduzione, database bilingui, si compongono di frasi nella lingua di partenza e le corrispondenti traduzioni nella lingua di arrivo, e possono proporre frasi intere o segmenti come corrispondenze, o *matches*. L'utente può inserire anche informazioni aggiuntive: meta-informazioni come data di inserimento, autore e informazioni personalizzate, ad esempio si può specificare che una frase è stata tradotta in un determinato modo per uno specifico cliente. Oltre ai *matches* presentati nel paragrafo 1.2.1, Trados Studio presenta anche due altri tipi: *context match* e *perfect match*. Il *context match* si verifica quando i contesti del segmento da tradurre e del segmento memorizzato sono gli stessi. Per determinarlo, il sistema considera informazioni aggiuntive, quali la struttura del testo e la progressione dei segmenti. Come osservano Lecci e Di Bello (2012: 26), nel caso di un *context match* viene valutata non solo la funzione del segmento memorizzato e di quello da tradurre, ma anche la corrispondenza tra il segmento precedente del testo di partenza e quello del segmento memorizzato. Il *perfect match*, invece, consiste in un confronto a livello di documenti: il testo in traduzione viene confrontato con un documento già tradotto in precedenza, solitamente una versione precedente del testo che si sta traducendo. In questo caso, tutte le traduzioni dei segmenti uguali presenti nei due documenti vengono inserite automaticamente. Una volta salvati nella memoria di traduzione, i *matches* possono essere comunque modificati e salvati nuovamente.

Per quanto riguarda i database terminologici in Trados Studio, questi possono essere creati e gestiti con MultiTerm. Si tratta di un sistema di gestione della terminologia la cui licenza viene fornita unitamente a quella di Trados Studio. MultiTerm è a tutti gli effetti un tool autonomo e può essere usato sia con Trados Studio che in modo indipendente⁴⁴.

Un'ulteriore risorsa rilevante è costituita dai dizionari di AutoSuggest, database bilingui contenenti frammenti linguistici (parole, espressioni) in una lingua di partenza e i rispettivi equivalenti nella lingua di arrivo. Tali dizionari sono creati "a partire da grandi memorie di traduzione e vengono consultati automaticamente" (Lecci e Di Bello 2012: 26). Durante la digitazione, i dizionari di AutoSuggest suggeriscono automaticamente il frammento di testo che si intende inserire. I suggerimenti sono *context-sensitive*, ossia traducenti potenziali selezionati in base al contesto linguistico del segmento di partenza, e non come semplici equivalenti isolati. I dizionari di Autosuggest possono anche essere consultati manualmente, permettendo l'utente di esplorare direttamente le corrispondenze bilingui memorizzate.

⁴⁴ Nel paragrafo 1.2.1 è stato descritto il funzionamento dell'editor di testo di Trados Studio.

5.2.2.2 Come funziona Trados Studio?

Il testo che si intende tradurre può essere in uno dei tanti formati supportati dal CAT tool. Al momento dell'avvio del progetto di traduzione, il traduttore apre il documento nel tool e collega tutte le risorse disponibili, ad esempio una memoria di traduzione. Il programma genera automaticamente il file SDLXLIFF: si tratta del formato bilingue dei file di Trados Studio. Successivamente si può procedere con la traduzione. È importante ricordare che, una volta conclusa la traduzione di un segmento, questo deve essere confermato prima di procedere al segmento successivo. La conferma dei segmenti non impedisce successive modifiche, che possono essere effettuate ritornando semplicemente sul segmento interessato. Trados Studio offre anche diverse scorciatoie da tastiera per svolgere alcuni compiti: ad esempio, per confermare un segmento, l'utente può usare il comando *Ctrl + Invio*. Queste combinazioni di tasti servono a velocizzare il processo di traduzione, anche se all'inizio possono risultare difficili da memorizzare. Per questo motivo sono a disposizione dell'utente dei pulsanti nel *Work Pane*, presente nella parte superiore della schermata.

Trados Studio offre all'utente una serie di funzionalità. Una di queste è l'*Auto-propagation*: si tratta dell'“inserimento automatico della traduzione di un segmento source ripetuto all'interno dello stesso documento” (Lecci e Di Bello 2012: 37). Pertanto, se un segmento già tradotto viene ripetuto in seguito, l'utente non dovrà reinserire la traduzione, ma questa verrà inserita automaticamente dal CAT tool. È possibile anche generare concordanze di espressioni presenti nella memoria di traduzione. Questa funzione può risultare particolarmente utile per i casi in cui né la TM e né la TB (*Termbase*) offrano soluzioni traduttive (Lecci e Di Bello 2012: 27). Nel *Work Pane* è presente il bottone di “ricerca concordanze”, come mostra la Figura 12.

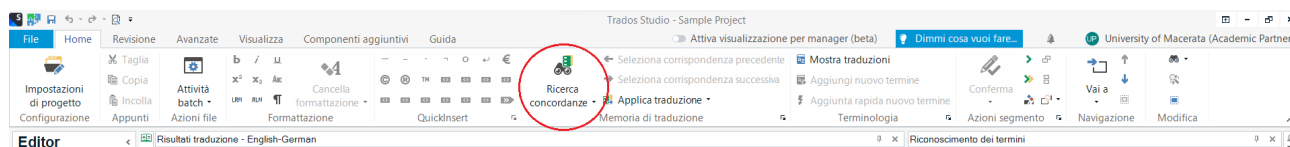


Figura 12: *Work Pane* di Trados Studio

Trados Studio consente inoltre di generare un'anteprima (*Preview*) del documento, che può rivelarsi estremamente utile, poiché consente all'utente di controllare l'andamento della traduzione e visualizzare l'aspetto finale del testo. Un'ulteriore funzione offerta da Trados Studio è il *QA Checker* che “consente di effettuare controlli a livello terminologico e formale (di formattazione, punteggiatura, ecc.), sia in tempo reale sia alla fine del processo di traduzione” (Lecci e Di Bello

2012: 27). Questa funzione permette di individuare errori o incongruenze nei segmenti tradotti, garantendo maggiore coerenza e accuratezza terminologica.

Una volta ultimata la traduzione, la memoria di traduzione può essere aggiornata definitivamente e la traduzione deve essere salvata. Infine, l'utente ottiene un file nello stesso formato del testo di partenza, pronto per la consegna finale al cliente.

Come già menzionato precedentemente, è possibile integrare nei CAT tool dei sistemi di TA. Per Trados Studio RWS offre Language Weaver. È possibile integrare anche sistemi di TA esterni, come nel presente studio è stato fatto con DeepL Pro. Il sistema di TA integrato proporrà la traduzione nel momento in cui non ci saranno suggerimenti da parte della memoria di traduzione (Lecci e Di Bello 2012: 26).

5.2.3 Sketch Engine

Sketch Engine è uno strumento online a pagamento, sviluppato nel 2003 da Lexical Computing Limited, progettato per la gestione di corpora e l'analisi testuale. Attualmente supporta più di 90 lingue e mette a disposizione degli utenti circa 600 corpora pronti all'uso. Consente anche la creazione di corpora personalizzati, nei quali l'utente può specificare informazioni e caratteristiche come il nome del corpus, la tipologia mono- o plurilingue, la lingua di riferimento e, facoltativamente, può aggiungere una breve descrizione. Per creare un corpus è possibile utilizzare testi reperiti dal web oppure selezionare file già in proprio possesso. È possibile creare anche subcorpora.

Sketch Engine offre varie funzionalità che lo rendono particolarmente utile in ambito linguistico e terminologico. La ricerca con questo tool su uno o più corpora consente di ottenere in modo automatizzato diversi tipi di informazione in base a quello di cui si ha bisogno. Non tutte le funzioni sono sempre disponibili e questo può dipendere dal corpus e/o dall'abbonamento che si possiede. Le funzioni utilizzabili sono evidenziate in bianco, mentre quelle non utilizzabili restano in azzurro, come si può osservare nella dashboard illustrata nella Figura 13.

DASHBOARD

Italian Web 2020 (itTenTen20)



ITALIAN WEB 2020 (ITTENTEN20)
INFORMAZIONI SUL CORPUS
GESTISCI CORPUS

Word Sketch
Collocazioni e combinazioni lessicali

Differenze tra Word Sketch
Confronta collocazioni di due parole

Thesaurus
Sinonimi e parole simili

Concordanze
Esempi d'uso in contesto

Concordanze Parallele
Ricerca traduzione

Lista di parole
Lista di frequenza

N-grammi
Espressioni multi-parola (Multi Word Expressions)

Parole chiave
Estrazione di terminologia

Trends
Analisi diacronica, neologismi

Analisi tipi di testo
Statistiche dell'intero corpus

OneClick Dictionary
Stesura automatica di dizionario

Termini bilingue
Estrazione di terminologia bilingue

Figura 13: *Dashboard* di Sketch Engine

La funzione più nota è *Concordanze*, che consente di estrarre concordanze KWIC (*key words in context*), cioè occorrenze di una parola o espressione nel corpus, mostrando gli esempi nel loro contesto d'uso. La ricerca può essere condotta in due forme: *Base* e *Avanzate*. Con la ricerca *Base*, si digita una parola o sintagma (in forma lemmatica o no) e cliccando su “Cerca” si ottengono tutte le occorrenze di quella data parola o sintagma all'interno del corpus, nel loro relativo contesto di occorrenza (cfr. Fig. 14).

CONCORDANZE
Corpus_IT_OFF

semplice ricerca • 1.235
1.039,2 token per milione • 0.1%

Ottieni più spazio

KWIC

☐ Dettagli
Contesto sinistro
KWIC
Contesto destro

1	☐	doc#0	centi, non è sorta dal nulla, ma ha progressivamente implementato il suo oggetto di ricerca e le sue ipotesi metodologiche anche mediante l'interazione con discipline già affer
2	☐	doc#0	nulate a proposito della terminologia adoperata: È già da tempo che fra gli ambiti di ricerca coltivati rientra la terminologia linguistica, in specie quella traduttologica; per un apri
3	☐	doc#0	con- cetti e termini e dall'altro l'influsso del punto di vista del curatore nell'ambito di ricerca nella traduzione, rassicura il lettore facendo notare come le "most cur- rent perspec
4	☐	doc#0	><s> Tale interrelazione tra concetto/termine e ambito d'uso, nonché orizzonte della ricerca , si rivela di indubbia utilità anche dal punto di vista didattico.</s><s> Spostando l'al
5	☐	doc#0	nettendo laconicamente la difficile delimitazione e presentazione del loro oggetto di ricerca , overossia la traduzione, riconoscono di essere partiti da obiettivi ambiziosi: Nonc
6	☐	doc#0	5 La presenza di tali risorse quale imprescindibile sostegno per la didattica e per la ricerca garantisce all'attività traduttiva di imporsi sempre più come settore di indagine, il cu
7	☐	doc#0	i proble- matici in parte da attribuire alla recenziarietà della disciplina, i cui indirizzi di ricerca si trovano ancora in una fase di intenso sviluppo.</s><s> Tale condizione si riflette i
8	☐	doc#0	, di cui qui è impossibile discutere la validità, si inseriscono appieno in quel filone di ricerca volto ad analizzare l'assetto lin- guistico delle traduzioni diffuse in italiano.</s><s> L
9	☐	doc#0	e sincronica applicata alla terminologia italiana della traduzione apre uno spazio di ricerca fino a ora esplorato solo parzialmente.</s><s> Per un'ulteriore puntualizzazione si l
10	☐	doc#1	la verifica della portata della "clausola di salvaguardia" si correla, di necessità, alla ricerca e alla ricostruzione dell'assetto di interessi disciplinato dal diritto positivo.</s><s> Al
11	☐	doc#2	otto controllo e gestire le proprie attività in maniera proficua.</s><s> L'oggetto della ricerca o della configurazione del campo da gioco La traduzione come attività economica l
12	☐	doc#2	resto il senso della specializzazione.</s><s> Ciò comporta da parte del singolo una ricerca abbastanza laboriosa che, partendo dalle capacità personali, si sviluppa in un'espe
13	☐	doc#2	ivo come quello della traduzione finanziaria – che richiede concentrazione, studio e ricerca – è importante avere il massimo della tranquillità.</s><s> Per questo è importante s
14	☐	doc#2	ci perché le situazioni cambiano vuoi perché il traduttore, nel suo costante lavoro di ricerca , scopre soluzioni linguistiche migliori rispetto a quelle utilizzate fino a quel momen
15	☐	doc#3	erbi procomplementari nella prosa letteraria tra diacronia e traduzione Ambito della ricerca , materiali e metodi Questo contributo rende conto dei risultati parziali di una ricer
16	☐	doc#3	ricerca, materiali e metodi Questo contributo rende conto dei risultati parziali di una ricerca che ho intra- preso da tempo (Ondelli 2017 e Ondelli 2018) e che è tutt'ora in via di

Figura 14: Funzione *Concordanze* per la parola “ricerca”

Come si può notare nella Figura 14, Sketch Engine evidenzia le occorrenze in rosso e mostra il contesto che le precede (*left context*) e che le segue (*right context*). In alto a sinistra è possibile vedere anche il numero di occorrenze individuate nel corpus dal tool.

È possibile eseguire diversi tipi di ricerche, ad esempio:

- cercare una forma lemmatica per ottenere tutte le occorrenze di quella parola in tutte le sue varianti (ad esempio analizzare -> analizza, analizzati, analizzando ecc.);
- cercare una forma non lemmatica ottenendo esempi in cui compare esattamente la forma scelta (es. analizzati -> analizzati);
- inserire un asterisco tra le parole oppure all'inizio/fine di una parte della parola; in questo modo l'asterisco assume la forma di un numero, di una parola o di un qualsiasi carattere generico (ad es. esamin* -> esamina, esaminano, esaminate ecc.; un * scientifico -> un articolo scientifico, un testo scientifico, un discorso scientifico);
- si possono inserire uno o più punti interrogativi ponendoli, ad esempio, all'inizio o alla fine di una parola e questi assumeranno la forma di un singolo carattere generico (es. ****ato -> evocato, passato, coniato: si ricercano parole o lemmi di 7 lettere che terminano per -ato).

Nella versione *Avanzate* di Concordanze, Sketch Engine offre varie opzioni di ricerca:

- semplice: si effettua lo stesso tipo di ricerca della versione Base; quindi, si possono digitare singole parole o interi sintagmi e, se scritti in forma lemmatica, Sketch Engine troverà tutte le varie forme attestate nel corpus;
- lemma: vengono cercate tutte le forme appartenenti al lemma che si indica. È possibile anche specificare quale parte del discorso si intende cercare (ad es. nomi, verbi, aggettivi, ecc.) selezionandolo dalla tabella "Parte del discorso" accanto all'elenco "Tipo di query".
- sintagma: si ricerca esattamente il sintagma digitato (ad es. condurre una ricerca -> condurre una ricerca);
- parola: si otterranno dei risultati esattamente per la parola indicata (ad es. risultati -> risultati);
- carattere: vengono cercate tutte le parole o *tokens* che contengono il carattere o la sequenza di caratteri di interesse (ad esempio mente -> relativamente, parallelamente);
- CQL: la *Corpus Query Language* è utile per effettuare ricerche complesse: è possibile combinare tutte le opzioni di ricerca illustrate in precedenza in un'unica ricerca che assume la forma di un'equazione, in cui ogni *token* ricercato viene espresso secondo regole precise all'interno di parentesi quadre ([]). Questo tipo di ricerca consente di trovare schemi lessicali

o grammaticali senza specificare parole concrete (es. tutti i nomi non preceduti da un articolo ma seguiti da un verbo al tempo passato).

Un'ulteriore funzione offerta da Sketch Engine è *Word Sketch*, che consente di osservare le collocazioni di una parola, ovvero le combinazioni ricorrenti che essa forma con altre parole. Si digita il lemma e cliccando su “VAI” si ottiene un elenco delle collocazioni del lemma indicato all'interno del corpus selezionato (cfr. Fig. 15).

verbs with "analisi" as object	verbs with "analisi" as subject	modifiers of "analisi"	noun modifiers of "analisi"	prepositional phrases with nouns	pronominal possessors of "analisi"
condurre condurre un' analisi	comparare che l' analisi comparata	quantitativo analisi quantitativa	fattoriale L' analisi fattoriale ci	"analisi" del "analisi" della	nostra la nostra analisi
effettuare effettuare analisi	approfondire analisi approfondita	contrastiva un' analisi contrastiva	corpus-based analisi corpus-based	"analisi" di "analisi" per	sua nella sua analisi
permettere permettere un' analisi	permettere analisi permette di	critico un' analisi critica	dell analisi dell	"analisi" sul "analisi" al	loro nella loro analisi
proporre propone un' analisi	incrociare analisi incrociata	qualitativo analisi qualitativa		"analisi" a "analisi" con	
estendere estendere le analisi	basire analisi si basano	statistico di analisi statistica			
basare basata l' analisi	evidenziare analisi ha evidenziato	comparativo analisi comparativa			
concentrare concentrata l' analisi	basare analisi basata	testuale di analisi testuale			
svolgere svolto una breve analisi	prendere analisi non ha preso	stilistico L' analisi stilistica			
compiere compie un' analisi	potere analisi può	attento un' attenta analisi del testo originale			
richiedere applicabili alla narrativa analisi		psicometriche analisi psicometriche con la IRT			

Figura 15: Funzione *Word Sketch* per la parola “analisi”

Come illustra la Figura 15, le concordanze vengono mostrate suddivise per classe grammaticale, ordinandole dalle più frequenti alle meno frequenti. Inoltre, in automatico viene fornita, sotto ogni parola mostrata, la rappresentazione più frequente della collocazione, riportando un esempio tratto dal corpus preso in esame.

Oltre alla funzione *Word Sketch*, Sketch Engine offre *Differenze tra Word Sketch* che può essere definita come un'estensione, poiché consente di confrontare due parole e il loro utilizzo. Esiste inoltre la funzione *Thesaurus* utile per la ricerca di sinonimi di una data parola. Digitando il lemma, l'utente accede a un elenco di tutti i sinonimi di quella data parola all'interno del corpus selezionato con le rispettive frequenze (cfr. Fig. 16).

	Lempos	Frequenza ?
1	dato	732 ...
2	esito	103 ...
3	obiettivo	700 ...
4	caso	1.838 ...
5	problema	571 ...
6	fase	698 ...
7	domanda	528 ...
8	prodotto	373 ...
9	proposta	552 ...
10	scopo	376 ...

Righe per pagina: 10 ▼ 1-10 di 219 < > 1 / 22 > >|

Figura 16: Funzione *Thesaurus* per la parola “risultato”

La funzione *Parole chiave* consente l'estrazione terminologica confrontando il corpus d'interesse con un corpus di riferimento. È possibile estrarre sia parole singole che espressioni multiparola. L'ultima funzione presentata in questa sede è *Concordanze parallele*, mediante la quale è possibile effettuare ricerche su due o più corpora paralleli trovando una parola o un sintagma nel primo corpus, e quindi nella prima lingua, e i corrispondenti segmenti tradotti nell'altro o negli altri corpora, e quindi nell'altra o nelle altre lingue. L'ampiezza delle funzioni disponibili rende Sketch Engine un ambiente versatile per l'analisi linguistica, particolarmente adatto a ricerche lessicali e terminologiche.

Nel presente progetto, Sketch Engine è stato impiegato per analizzare il corpus in lingua italiana e, durante la fase di *post-editing*, per estrarre parole ed espressioni tipiche del linguaggio scientifico. Questo ha consentito di identificare pattern ricorrenti e strutture lessicali caratteristiche del testo scientifico, fornendo un supporto concreto nella fase di correzione dell'output della TA e alla valutazione della qualità della traduzione generata automaticamente.

5.3 Fasi dello studio

Lo studio empirico si è articolato in diversi stadi operativi. In primo luogo, a seguito della selezione dei materiali e degli strumenti, i due corpora sono stati impostati sulle piattaforme prestabilite a seconda dell'uso e della finalità: il corpus in lingua italiana su Sketch Engine, il corpus in lingua tedesca su Trados Studio. Successivamente, sono state definite e condotte le diverse fasi, a partire dalla creazione dei progetti di traduzione all'interno di Trados Studio fino alla fase conclusiva di *post-editing*. L'intero percorso ha condotto all'elaborazione di linee guida specifiche, progettate per

la traduzione del linguaggio scientifico nell'ambito linguistico, con particolare riferimento alla coppia linguistica tedesco-italiano.

5.3.1 Creazione dei primi progetti su Trados Studio

Nella fase iniziale si è proceduto con la configurazione del corpus di lingua tedesca in progetti di traduzione sul tool Trados Studio. Sono stati creati tre progetti distinti, corrispondenti alle tre riviste da cui sono stati estrapolati gli articoli, al fine di agevolare la gestione del lavoro e, al contempo, osservare eventuali ricorrenze di errori traduttivi in relazione all'argomento trattato e/o alla rivista di provenienza. Nella fase di creazione dei progetti di traduzione, è stato integrato il sistema di TA DeepL Pro, permettendo la visualizzazione della traduzione generata dal sistema di TA direttamente nell'editor di testo del tool. Per ciascun progetto sono stati inoltre predisposti una memoria di traduzione e un database terminologico dedicato, strumenti utili rispettivamente a raccogliere e riutilizzare i segmenti tradotti e a garantire coerenza terminologica nel corso di tutto il processo traduttivo.

Durante la configurazione del progetto, nella schermata di "Impostazioni del progetto" sono state selezionate diverse opzioni volte a ottimizzare il flusso di lavoro. Nella schermata 2 è stata attivata l'opzione che permette di modificare i segmenti di partenza, rendendo possibile intervenire direttamente sul testo di partenza, qualora necessario. Inoltre, nella schermata 8, tramite il pulsante "Tutte le impostazioni di progetto", è possibile accedere alla sezione "Strumenti di controllo della qualità 3.0", nella quale sono stati selezionati parametri specifici per il controllo automatico di errori ricorrenti, come ad esempio la presenza di spazi in eccesso, al fine di garantire un livello di qualità più elevato e omogeneo nelle traduzioni prodotte (cfr. Fig. 17).

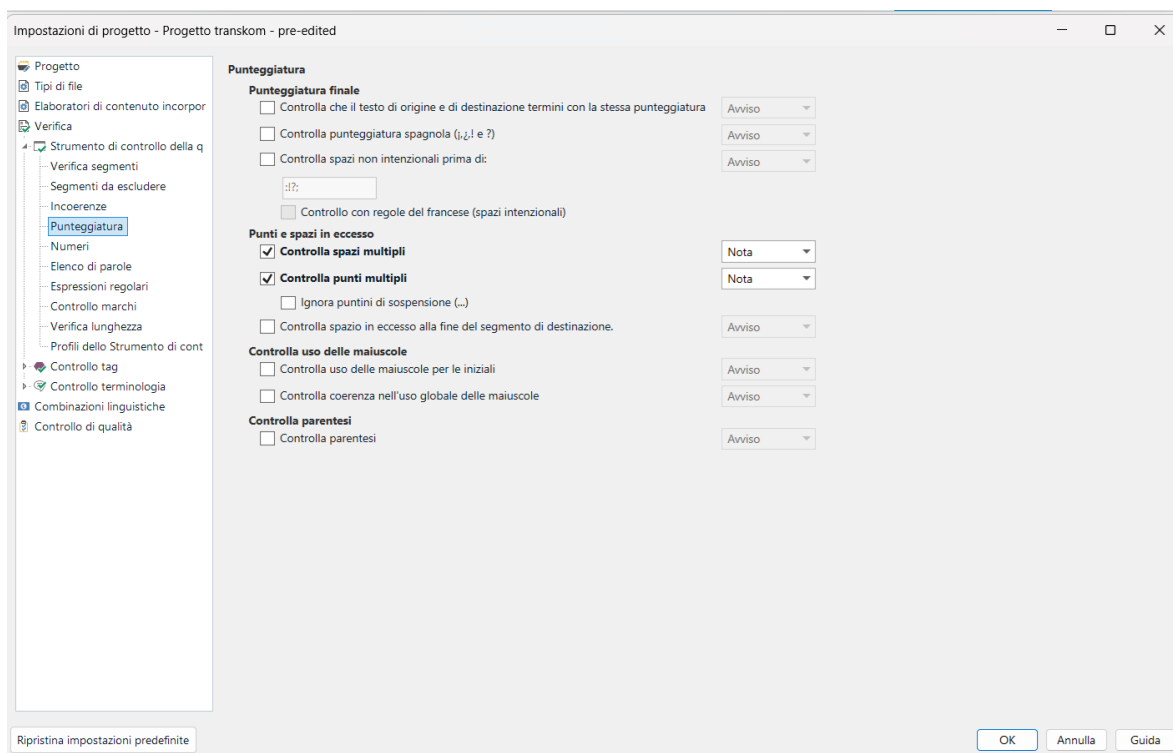


Figura 17: Strumenti di controllo della qualità 3.0

I file PDF degli articoli in lingua tedesca sono stati caricati su Trados Studio e successivamente convertiti in formati traducibili direttamente dal CAT tool. Una prima fase di analisi degli output della TA è stata eseguita direttamente sui file Word esportati da Trados Studio e contenenti l'output grezzo generato dal sistema. Per ciascun testo di un progetto, Trados Studio consente la creazione di un file Word scaricabile tramite la funzione “finalizza”, utilizzata solitamente a conclusione del lavoro di traduzione per ottenere il documento da consegnare al cliente. In questo caso si è “finalizzata” la traduzione senza averla conclusa realmente. Non è quindi stata apportata alcuna modifica al testo generato dal sistema, così da ottenere un file Word con l'output grezzo. Successivamente, i file Word sono stati analizzati per individuare gli errori presenti nella traduzione automatica. Gli errori sono stati evidenziati mediante una codifica cromatica: rosso per gli errori grammaticali, giallo per gli errori di senso, verde per gli errori di formulazione e celeste per gli errori in dubbio e da ricontrollare.

Durante questa fase preliminare sono emerse alcune problematiche legate al formato dei file. In primo luogo, le frasi che si estendono su due pagine, cioè quelle che iniziano in una pagina e si concludono nella successiva, venivano spezzate e restituite nell'output della TA come due frasi, una che finiva in corrispondenza della fine della prima pagina e un'altra nuova che iniziava in corrispondenza dell'inizio della pagina successiva. Nella gran parte dei casi entrambe le frasi venivano tradotte, ma risultavano prive di coerenza tra loro, compromettendo il senso complessivo

del testo. Si sono verificati anche casi in cui il sistema di TA tentava di completare la prima parte della frase indovinandone la conclusione e ignorando la seconda parte; o ancora dei casi in cui la prima parte della frase veniva spezzata e tradotta con un punto finale, mentre la seconda parte rimaneva nella lingua di partenza (cfr. Fig. 18).

La denominazione viene spesso utilizzata per registrare le competenze lessicali.
Rispondere alle domande e completare le frasi sono metodi utilizzati internamente.

Indagine sulle competenze sintattiche dei bambini — 427

halb der elizitierten Produktion, die es auch ermöglichen, dem Kind syntaktische
Strukturen zu „entlocken“. Al contrario, le descrizioni di singole immagini o di
storie di immagini all'interno del processo di elicitazione sono formati aperti.

Figura 18: Esempio di resa problematica con Trados Studio

Il titolo dei contributi, riportato nell'intestazione ripetuta all'inizio di ogni pagina, presentava diversi problemi di traduzione: in alcuni casi non veniva tradotto, in altri veniva tradotto solo parzialmente o in modo non uniforme; talvolta compariva in inglese o risultava diverso rispetto alla traduzione proposta sulla prima pagina. Analogamente, quando l'intestazione riportava informazioni relative alla rivista, quali titolo, volume o edizione, queste non venivano restituite correttamente. In numerosi casi, il titolo della rivista veniva tradotto, pur non essendo necessario, oppure venivano tradotti solo alcuni elementi, generando incoerenze.

Un'ulteriore problematica riscontrata riguarda le parole spezzate dal trattino breve a fine riga. In numerosi casi, il sistema di TA ha preservato il trattino anche nell'output, riportando la parola in modo scorretto. Inoltre, durante la conversione del testo di partenza in un file traducibile, si sono riscontrate anomalie di spaziatura tra le parole: in alcuni casi gli spazi erano eccessivi, in altri completamente assenti. Queste criticità presenti nel testo di partenza sono state riportate dal sistema di TA anche nel testo di arrivo.

Alla luce delle problematiche riscontrate, con l'intento di concentrare l'analisi esclusivamente sulla resa del linguaggio scientifico da parte del sistema di TA ed eliminare quindi gli errori ricorrenti non strettamente collegati alla traduzione del linguaggio scientifico, si è deciso di operare interventi di *pre-editing* descritti di seguito.

5.3.2 Interventi di *pre-editing*

Il termine “pre-editing”, già definito precedentemente nel paragrafo 1.1., indica interventi volti a preparare il testo alla traduzione automatica, al fine di ottimizzare le prestazioni del sistema di TA scelto.

Nel presente studio, sono stati eseguiti i seguenti interventi di *pre-editing*:

- eliminazione della spaziatura superflua tra parole;
- aggiunta della spaziatura mancante tra parole;
- rimozione dei trattini brevi utilizzati per dividere le parole a fine riga;
- eliminazione delle intestazioni riportanti il numero della pagina o informazioni sull’articolo e/o rivista (es. titolo dell’articolo, titolo della rivista, edizione, volume ecc.);
- ricongiunzione delle frasi che si estendono su due pagine.

In merito a quest’ultimo intervento, come precedentemente illustrato, si tratta di frasi la cui estensione si articola su due pagine consecutive. In fase di *pre-editing* è emerso che tali frasi risultavano spezzate verosimilmente a causa di un’interruzione di sezione inserita a fine pagina nel documento originale. Tale problematica sembra riconducibile al formato originale del testo di partenza, un PDF, convertito successivamente in file Word dal CAT tool. Tale “caratteristica”, pur essendo un dettaglio di layout, genera un errore traduttivo che viene riportato dal sistema di TA nel testo di arrivo.

Un ulteriore intervento rilevante di *pre-editing* riguarda gli articoli estratti dalla rivista *Interculture journal*, in cui il testo si presenta diviso su due colonne (cfr. Fig. 19).



An- derer berücksichtigt; kulturelle Un- terschiede werden bearbeitbar gemacht, indem die professionelle Begleitperson, hier der Coach, kulturkreisspezifisches Wissen nutzt, zur Verfügung stellt oder bei ihren Deutungen bzw. Diagnosen darauf zurückgreift und so dessen An- wendung praktisch demonstriert.	Schroll-Machl 2006:20 und Steixner 2009).
In diesem Sinne begreifen manche Definitionen interkulturelles Coaching als eine exklusive Personalentwick- lungsmaßnahme (Ostermann 2001, Barmeyer 2003, Kinast 2003) und Form interkulturellen Lernens mit dem Ef- fekt, dass bestimmte Ziele, Inhalte und Interventionen – wie bei einem Trai- ning – bereits vorgegeben sind oder sich zumindest stark aufdrängen. „Coaching kann helfen, kulturell bedingte Irritatio- nen und Störungen („Critical Incidents“) konstruktiv zu thematisieren und durch prozessual begleitende Anstöße den Coach zur Selbstreflexion und zum Finden von Lösungsstrategien anzure- gen“ (Barmeyer 2003:21).	Die zugrunde liegende Kulturvorstel- lung kann als traditionell bezeichnet und bis auf das Kugelmodell von Her- der zurückverfolgt werden. Kultur wird als ein homogenes System gefasst, d. h. als eine in sich geschlossene und nach außen klar abgrenzbare Lebensweise bzw. als eine mit nationalstaatlichen Grenzen deckungsgleiche Totalität. Diese Auffassung betont den essentiali- stischen und statischen Kern von Kultur und fokussiert die Kohärenz zwischen Wohngebiet, Wirtschaft und geteilten Normen und Werten in Abgrenzung zu anderen <i>Kulturcontainern</i> , deren Mitglieder als <i>Träger</i> von Kultur mit entsprechenden Prägungen aufgefasst werden. Nach diesem Kulturkonzept tauchen die Coachingpartner in eine für sie fremde Kultur ein oder es begegnen sich Repräsentanten verschiedener Kulturen. Dabei entsteht eine inter- kulturelle „Überschneidungssituation“

Figura 19: Esempio n.1 *pre-editing*

A seguito delle problematiche riscontrate legate alla disposizione del testo in colonne, che comportava anche un uso massiccio del trattino (-) per segmentare le parole a fine riga, si è deciso in fase di *pre-editing* di riorganizzare il testo in modo più lineare, disponendolo su un'unica colonna. Oltre a questo intervento specifico, sono stati eseguiti ulteriori interventi di *pre-editing* analoghi a quelli eseguiti per gli articoli delle altre due riviste: rimozione dei capoversi che interrompevano le frasi e rimozione dei trattini tra le parole (cfr. Fig. 20).

kulturkreisspezifisches Wissen nutzt, zur Verfügung stellt oder bei ihren Deutungen bzw. Diagnosen darauf zurückgreift und so dessen Anwendung praktisch demonstriert.

Figura 20: Esempio n.2 *pre-editing*

5.3.3 Creazione dei secondi progetti su Trados Studio

In seguito agli interventi di *pre-editing*, è stata ripetuta la procedura illustrata nel paragrafo 5.3.1. Sono stati pertanto creati su Trados Studio tre progetti, ciascuno corrispondente alle tre riviste da cui provengono gli articoli scientifici analizzati nello studio. Al momento della creazione di ciascun progetto, sono stati caricati i file Word contenenti gli articoli scientifici pre-editati. Il sistema di TA DeepL è stato nuovamente integrato, consentendo di visualizzare la traduzione generata automaticamente direttamente nell'editor di testo nel momento in cui si apre uno dei file del progetto.

Nonostante le modifiche apportate in fase di *pre-editing*, è stato necessario intervenire ulteriormente sul testo di partenza. Il CAT tool in questione permette, infatti, di apportare modifiche sul testo di partenza mediante specifiche opzioni selezionabili al momento della creazione del progetto. L'intervento necessario riguardante la segmentazione del testo di partenza prevede l'unione di due segmenti nel caso in cui il primo segmento finisca con i due punti (:), come mostrano le figure 21 e 22.

32 So geben sie Kritikern von Übersetzung das Rüstzeug an die Hand, von dem sie fälschlich glauben, einen Sinnbezug von Texten (oder gar von Text an sich) mit relativistischen Behauptungen destruieren zu können.¹⁰ Wenngleich mit anderer Stoßrichtung, aber in der Konsequenz vergleichbar schreibt Erich Prunč zur Skopostheorie in <i>Entwicklungslinien der Translationswissenschaft</i>. 33 Der Skopos einer kommunikativ adäquaten Translation ist in der konfliktträchtigen multikulturellen Gesellschaft des beginnenden 21. Jahrhunderts ohne Zweifel die sinnvollste Vorgabe für translatorisches Handeln.	NMT In questo modo, forniscono ai critici della traduzione gli strumenti che erroneamente credono di poter utilizzare per distruggere il significato dei testi (o addirittura del testo stesso) con affermazioni relativistiche.¹⁰ Anche se con un'impostazione diversa, ma con conseguenze paragonabili, Erich Prunč scrive sulla teoria scopica in <i>Linee di sviluppo di Studi sulla traduzione</i>. NMT Nella società multiculturale e conflittuale dell'inizio del XXI secolo, lo skopos di una traduzione comunicativamente adeguata è senza dubbio la linea guida più sensata per l'azione traduttiva.

Figura 21: Esempio n.1 di errore di segmentazione

121 Bei seiner Suche nach dem Prinzip, das eine Gesellschaft, in der viele verschiedene Teile, die beileibe nicht alle miteinander kongruent oder füreinander verständlich sind, zu einer Gesellschaft macht, kommt er zu dem Phänomen der Übersetzung. 122 Die Suche nach der "Einheit" der Gesellschaft und nach ihrer "Konstitution" konkretisiert sich zu der Erkundung von praktischen Formen, Grenzen zwischen Teilen der Gesellschaft zu überschreiten, ohne sie aufzulösen oder aufzuheben.	NMT Nella sua ricerca del principio che fa di una società in cui sono presenti molte parti diverse, che non sono affatto tutte congruenti o comprensibili tra loro, una società, egli giunge al fenomeno della traduzione. NMT La ricerca dell'"unità" della società e della sua "costituzione" si concretizza nell'esplorazione di forme pratiche per superare i confini tra le parti della società senza dissolverli o abolirli.

Figura 22: Esempio n.2 di errore di segmentazione

Questo intervento si è rivelato necessario in quanto la segmentazione eseguita da Trados Studio comporta la formazione di due segmenti distinti: il primo termina con i due punti, mentre il secondo inizia con una lettera maiuscola. Tale punteggiatura viene riprodotta fedelmente anche nel testo di arrivo, generando un errore grammaticale in italiano, dove la lettera maiuscola dopo i due punti non è necessaria. Unendo i due segmenti, il sistema di TA DeepL corregge automaticamente questo errore, adoperando nel segmento in italiano la lettera minuscola dopo i due punti, come si può osservare nella figura 23.

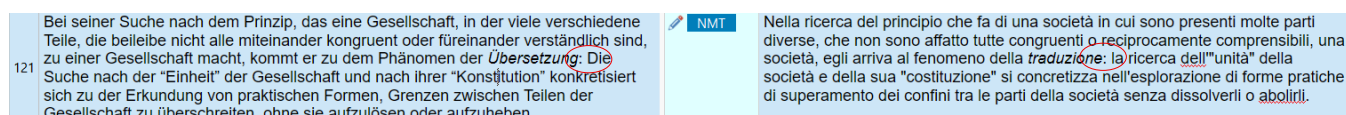


Figura 23: Esempio n.1 segmentazione corretta

Dopo aver effettuato questi ulteriori interventi necessari, le traduzioni sono state finalizzate, ottenendo per ciascun file un documento Word contenente l'output grezzo.

5.3.4 Individuazione e annotazione degli errori

La fase successiva ha riguardato l'individuazione degli errori presenti nei contributi tradotti automaticamente. L'analisi è stata condotta in modo sistematico, prendendo in esame i testi singolarmente. In una prima fase esplorativa, sono stati selezionati cinque testi per ciascuna fonte, per un totale di quindici testi, sui quali è stata eseguita un'analisi preliminare volta a individuare i principali errori ricorrenti. Gli errori riscontrati sono stati provvisoriamente ricondotti a tre macrocategorie:

- S (errori semantici), evidenziati in giallo;
- G (errori grammaticali), evidenziati in rosso;
- F (errori di formulazione), evidenziati in verde.

L'uso di colori diversi si è rivelato uno strumento pratico per visualizzare e distinguere rapidamente i diversi tipi di errore nei file Word. A ciascun errore individuato si è assegnato un nome, una parola chiave che ne descrivesse la natura, così da impostare una classificazione più precisa. Questa fase di lavoro è stata guidata da alcuni modelli di riferimento, linee guida, ampiamente utilizzati negli studi sulla qualità della traduzione e del *post-editing* che verranno descritti dettagliatamente in seguito (cfr. 5.3.5).

Conclusa l'analisi preliminare, che ha permesso di raccogliere i dati iniziali e riflettere sui criteri di classificazione, è stata avviata la fase di sistematizzazione. L'attività di annotazione è stata temporaneamente interrotta per elaborare una prima stesura di linee guida specifiche.

Successivamente, l'annotazione è stata ripresa con l'obiettivo di analizzare l'intero corpus secondo questa nuova classificazione degli errori. Durante questo secondo intervento annotativo, a ogni errore rilevato è stata assegnata una categoria specifica, annotata direttamente nella sezione "Commenti" dei file Word, così da garantire una documentazione chiara e coerente del processo. Parallelamente, le linee guida specifiche sono state progressivamente affinate e definite man mano che il lavoro di annotazione procedeva. L'individuazione degli errori pertanto non solo ha consentito di mappare le criticità ricorrenti nelle traduzioni automatiche, ma ha anche fornito la base empirica per l'elaborazione delle linee guida di classificazione degli errori specifiche per la tipologia di testi analizzati e descritte nel paragrafo successivo.

5.3.5 Tassonomia per Errori nella Traduzione Automatica Scientifica Tedesco-Italiana: TETASTI

Le linee guida TETASTI, sviluppate nel presente elaborato per la classificazione, la categorizzazione e l'annotazione degli errori nelle traduzioni automatiche, costituiscono uno strumento originale di valutazione della qualità traduttiva di articoli scientifici in ambito linguistico dal tedesco all'italiano. Esse sono state elaborate a partire dall'analisi empirica del corpus tedesco, tradotto con il sistema di TA DeepL, e costruite progressivamente in base all'individuazione e alla sistematizzazione degli errori ricorrenti. L'obiettivo è la realizzazione di uno strumento metodologico specifico per la coppia linguistica tedesco-italiano, per il genere testuale dell'articolo scientifico e per il dominio disciplinare della linguistica. Gli esempi riportati nelle linee guida provengono direttamente dai testi del corpus oggetto di studio, mentre le correzioni associate a ciascun caso sono state formulate manualmente sulla base dell'analisi linguistica e traduttiva condotta personalmente. Nei paragrafi successivi vengono presentati i modelli teorici e le linee guida già esistenti che hanno costituito il punto di partenza per la progettazione della tassonomia, per poi illustrare nel dettaglio la struttura e il funzionamento delle linee guida TETASTI messe a punto in questa ricerca.

5.3.5.1 Modelli di riferimento

Per la definizione delle linee guida elaborate nel presente progetto sono stati presi come riferimento i seguenti modelli consolidati di valutazione della qualità traduttiva, che verranno di seguito illustrate: *Korrekturrichtlinien dell'Istituto per Traduttori e Interpreti dell'Università di Heidelberg* (IÜD, Universität Heidelberg 1996), *Annotation Guidelines for English-Dutch Translation Quality Assessment, version 1.1* (Daems et al. 2013), *Annotation Guidelines for legal Italian<->(South Tyrolean) German Machine Translation Quality Assessment* (Chiocchetti e De Camillis 2022),

SCATE taxonomy (Tezcan et al. 2017) *Multidimensional Quality Metrics (MQM)* (Lommel, Uszkoreit, et al. 2014; Lommel et al. 2013).

- *Korrekturrichtlinien dell'Istituto di Traduzione e Interpretazione dell'Università di Heidelberg* (IÜD, Universität Heidelberg 1996)

Le linee guida di correzione (*Korrekturrichtlinien*) dell'Istituto di Traduzione e Interpretazione dell'Università di Heidelberg (IÜD) sono uno strumento didattico ampiamente usato nella formazione dei traduttori. Il modello è stato concepito per la valutazione accademica delle traduzioni, ma la sua chiarezza e praticità lo rendono funzionale anche in contesti di analisi della qualità traduttiva.

Queste linee guida classificano gli errori secondo tre macrocategorie: errori grammaticali (G) che riguardano la correttezza morfosintattica della lingua di arrivo (ad esempio concordanze, tempi verbali); errori semantici (S) che si riferiscono a scelte traduttive che compromettono la rappresentazione corretta del significato (ad esempio la traduzione scorretta di una parola, la perdita di una sfumatura semantica); errori di formulazione (F) che riguardano la resa comunicativa e funzionale del testo e quindi la sua naturalezza stilistica, l'adeguatezza al contesto, al destinatario e alla funzione del testo stesso (IÜD, Universität Heidelberg 1996). Con particolare riferimento alle macrocategorie G e S, gli errori possono emergere e quindi essere rilevati sia all'interno di una singola frase o di sue parti, sia oltre la singola frase, a livello transfrastico. Nel primo caso si tratta di problemi locali, come una concordanza errata o la resa sbagliata di un termine; nel secondo caso, invece, l'errore ha un impatto più ampio, compromettendo la coerenza del discorso o la continuità semantica del testo. A loro volta ogni macrocategoria presenta diverse sottocategorie.

Per garantire maggiore precisione nell'annotazione, gli errori vengono denominati attraverso un sistema codificato: la categoria principale è indicata con G, S o F, seguita da una barra e da un acronimo che specifica il tipo di errore. Ad esempio, *G/konj* segnala un errore nell'uso delle congiunzioni. Questo sistema permette di classificare gli errori in maniera chiara, sintetica e facilmente consultabile, rendendo più immediata l'analisi e il confronto tra testi diversi.

- *Annotation Guidelines for English-Dutch Translation Quality Assessment, version 1.1* (Daems et al. 2013)

Le *Annotation Guidelines for English-Dutch Translation Quality Assessment* si basano su una distinzione fondamentale tra due gruppi di categorie: *adequacy* (adeguatezza) e *acceptability* (accettabilità). L'adeguatezza riguarda la corrispondenza tra testo di partenza e testo di arrivo, ovvero la fedeltà con cui i contenuti del testo originale vengono trasferiti nella traduzione (Daems et

al. 2013). Le sottocategorie principali dell'accettabilità includono: *Grammar & syntax*, *Lexicon*, *Spelling & typos*, *Style & register* e *Coherence*. Ognuna di queste sottocategorie presenta ulteriori sottocategorie di errore. L'accettabilità, invece, si concentra esclusivamente sul testo tradotto, valutandone la conformità alle norme linguistiche e culturali della lingua di arrivo (Daems et al. 2013). I due gruppi di categorie presentano diverse sottocategorie di errore che possono essere ampliate, modificate o ridotte in base alle esigenze specifiche del progetto di valutazione. Gli errori di adeguatezza includono sottocategorie molto dettagliate, che permettono di individuare diverse tipologie di problemi: *Contradiction*, *Word sense disambiguation*, *Hyponymy*, *Hyperonymy*, *Terminology*, *Quantity*, *Time*, *Meaning shift caused by punctuation*, *Meaning shift caused by misplaced word*, *Deletion*, *Addition*, *Explicitation*, *Coherence*, *Inconsistent terminology* e *other meaning shift*.

Per il modello, Daems et al. (2013) propongono un approccio in due fasi: nella prima, gli annotatori esaminano soltanto il testo di arrivo e devono valutarlo in termini di accettabilità, senza essere influenzati dal confronto con il testo di partenza. Ciò consente di giudicare la coerenza interna e la naturalezza linguistica della traduzione. Nella seconda fase, gli annotatori ricevono il testo di partenza accostato alla traduzione e possono così concentrarsi sull'adeguatezza, verificando la fedeltà e la completezza del trasferimento dei contenuti.

- *Annotation Guidelines for legal Italian<->(South Tyrolean) German Machine Translation Quality Assessment* (Chiocchetti e De Camillis 2022)

Le *Annotation Guidelines for legal Italian<->(South Tyrolean) German Machine Translation Quality Assessment* si basano sulla versione 1.3.3 delle *Annotation Guidelines for English-Dutch Machine Translation Quality Assessment* e sono state sviluppate nell'ambito dello studio pilota MT@BZ presso l'Eurac Research di Bolzano. Queste linee guida presentano due macrocategorie di errori: errori di *accuracy* ed errori di *fluency*. Gli errori di *accuracy* riguardano la fedeltà del testo di arrivo rispetto a quello di partenza: rientrano in questa categoria tutte le occorrenze in cui il testo tradotto non riflette fedelmente le informazioni contenute nell'originale, dando luogo a fenomeni quali interpretazioni errate, contraddizioni, slittamenti di significato, aggiunte o omissioni (Chiocchetti e De Camillis 2022). Fanno parte della macrocategoria *accuracy* le seguenti sottocategorie di errori: *Addition*, *Omission*, *Untranslated*, *Do-Not-Translate* (DNT), *MisTranslation*, *Mechanical*, *Bilingual Terminology*, *Source Error* e *Other*. Le sottocategorie *MisTranslation* e *Mechanical* presentano al loro interno ulteriori categorie di errori (ad esempio *Semantically Unrelated* per *MisTranslation* e *Capitalization* per *Mechanical*). Gli errori di *fluency*, invece, sono legati esclusivamente alla costruzione linguistica del testo di arrivo e comprendono

tutte le deviazioni da norme e convenzioni della lingua di arrivo (Chiocchetti e De Camillis 2022). La macrocategoria *fluency* presenta le seguenti sottocategorie: *Grammar*, *Lexicon*, *Orthography*, *Multiple Errors*, *Coherence*, *Other*. Queste sottocategorie sono a loro volta suddivise in ulteriori sottogruppi di errori, ad accezione di *Multiple Errors* e *Other*.

Le linee guida forniscono inoltre indicazioni riguardanti l'annotazione mediante WebAnno, uno strumento che consente di associare ciascun errore individuato alla relativa categoria di errore, garantendo così uniformità e coerenza nel processo di annotazione.

- *SCATE taxonomy* (Tezcan et al. 2017)

La *SCATE taxonomy* nasce per la combinazione linguistica inglese-olandese e si basa su una netta distinzione tra due macrocategorie: *fluency* e *accuracy*. Gli errori di *fluency* riguardano la scorrevolezza del testo nella lingua di arrivo, indipendentemente dal trasferimento di significato dalla lingua di partenza. Si tratta dunque di errori che possono essere individuati analizzando il testo tradotto senza doverlo confrontare con l'originale (Tezcan et al. 2017). All'interno di questa categoria rientrano cinque sottogruppi principali: errori di grammatica (*grammar*), errori lessicali (*lexical choice*), errori ortografici (*orthography*), errori multipli (*multiple errors*) e altre tipologie di errori di *fluency* non classificabili nei gruppi precedenti (*other*). I sottogruppi *grammar*, *lexical choice*, *orthography* presentano rispettivamente un ulteriore livello di suddivisione contenente diverse categorie di errori. Gli errori di *accuracy*, invece, sono strettamente legati al contenuto: riguardano il grado di correttezza in cui il significato del testo di partenza viene rappresentato nel testo di arrivo (Tezcan et al. 2017). In questo caso, l'analisi richiede necessariamente il confronto tra il testo di partenza e il testo di arrivo, poiché si tratta di individuare discrepanze semantiche e distorsioni nella trasmissione del messaggio del testo di partenza. Gli errori di accuratezza vengono ulteriormente suddivisi in diverse categorie: errori relativi all'aggiunta di contenuti non presenti nel testo di partenza (*addition*), errori riguardanti l'omissione di contenuti presenti nell'originale (*omission*), segmenti lasciati nella lingua di partenza (*untranslated*), elementi che non avrebbero dovuto essere tradotti ma lo sono stati (*do-not-translate*), traduzioni errate (*MisTranslation*), errori legati a elementi formali o tecnici (*mechanical errors*), errori di terminologia bilingue (*bilingual terminology*), errori già presenti nel testo di partenza (*source errors*) e altre tipologie di errori di accuratezza non riconducibili a queste categorie. Le categorie *MisTranslation* e *mechanical errors* si compongono di un ulteriore livello di categorie di errori.

La *SCATE taxonomy* prevede l'annotazione degli errori di *fluency* solo nel testo di arrivo, mentre gli errori di *accuracy* vengono annotati sia nel testo di partenza che nel testo di arrivo,

collegando le due annotazioni in modo da evidenziare chiaramente la relazione tra errore e segmento originale.

- *Multidimensional Quality Metrics (MQM)* (Lommel, Uszkoreit, et al. 2014; Lommel et al. 2013)

Il modello MQM è uno dei più completi e versatili per la valutazione della qualità traduttiva ed è stato elaborato in versioni successive. Si propone come uno standard che può essere applicato alla valutazione della traduzione umana e automatica (Lommel, Uszkoreit, et al. 2014)⁴⁵. Esso presenta una tassonomia multilivello organizzata in dimensioni, ciascuna delle quali contiene categorie e sottocategorie di errore. Le principali dimensioni sono le seguenti⁴⁶:

- Terminologia (*Terminology*): errori legati all'uso scorretto della terminologia;
- Accuratezza (*Accuracy*): errori riguardanti la fedeltà al testo di partenza, ossia omissioni, aggiunte, significato errato;
- Convenzioni linguistiche (*Linguistic conventions*): errori relativi alla correttezza linguistica del testo, ossia errori grammaticali, ortografici, di punteggiatura e di formattazione di base. Nella versione 1 questa dimensione veniva indicata con il nome *Fluency*;
- Stile (*Style*): errori riguardanti l'adeguatezza stilistica, la naturalezza e il rispetto delle convenzioni stilistiche di genere testuale;
- Convenzioni locali (*Locale conventions*): errori che violano le convenzioni specifiche della lingua o della comunità di arrivo, inclusi aspetti di formattazione e rappresentazione di dati (ad es. date, unità di misura, valute);
- Adeguatezza al pubblico (*Audience appropriateness*): errori derivanti dall'uso di contenuti non validi o inadeguati per il pubblico o il contenuto di destinazione. Questa dimensione corrisponde alla dimensione *Verity* nella prima versione;
- Progettazione e markup (*Design and markup*): errori relativi alla progettazione grafica o alla presentazione del testo, inclusi errori di formattazione, struttura del layout, uso dei caratteri o integrazione con elementi grafici e interfacce;
- Personalizzazione (*Custom*): qualsiasi altro errore.

Questa organizzazione rende il modello MQM estremamente adattabile: le categorie possono infatti essere attivate, disattivate o modificate in base agli obiettivi del progetto di valutazione, consentendo un'analisi che tenga conto non solo degli aspetti traduttivi in senso stretto, ma anche delle esigenze del pubblico destinatario e della fruibilità del prodotto finale.

⁴⁵ Per maggiori informazioni si veda anche: <https://themqm.org/> (ultima consultazione: 20/12/2025)

⁴⁶ Per maggiori informazioni sulle macrocategorie e le rispettive sottocategorie, si veda: <https://themqm.org/the-mqm-typology/> (ultima consultazione: 20/12/2025)

In sintesi, i modelli analizzati condividono l'obiettivo di fornire criteri sistematici per la valutazione della qualità traduttiva, pur differendo per livello di dettaglio e ambito di applicazione. Le *Korrekturrichtlinien dell'Istituto per Traduttori e Interpreti dell'Università di Heidelberg* (IÜD, Universität Heidelberg 1996) offrono una classificazione essenziale e didatticamente efficace basata su tre macrocategorie (errori grammaticali, semantici e di formulazione). Le *Annotation Guidelines* di Daems et al. (2013) introducono la distinzione tra *adeguatezza* e *accettabilità*, ripresa e adattata anche nelle *Annotation Guidelines* di Chiocchetti e De Camillis (2022) e nella *SCATE taxonomy* di Tezcan et al. (2017), che ne ampliano la struttura e la precisione categoriale. Il modello *Multidimensional Quality Metrics* (Lommel, Uszkoreit, et al. 2014; Lommel et al. 2013) rappresenta invece un approccio più flessibile e modulare, articolato in diverse dimensioni, che consente di adattare la valutazione ai diversi contesti e obiettivi traduttivi. Nel complesso, tali modelli convergono su due assi fondamentali di analisi, *accuracy/adequacy* e *fluency/acceptability*, che costituiscono anche il riferimento teorico per la classificazione e l'analisi degli errori adottate nel presente studio.

5.3.5.2 Descrizione di TETASTI

Le linee guida TETASTI, qui elaborate, sono state sviluppate a partire dai summenzionati modelli di valutazione della qualità traduttiva, con l'obiettivo di creare uno strumento mirato per la coppia linguistica tedesco-italiano, il genere testuale dell'articolo scientifico e il dominio disciplinare della linguistica. Esse riprendono la distinzione tra *Accuracy* e *Fluency* presente nella *SCATE taxonomy* (Tezcan et al. 2017) e nelle *Annotation Guidelines* di Chiocchetti e De Camillis (2022), integrandola con la dimensione di *Grammar* introdotta dalle *Korrekturrichtlinien dell'Istituto per Traduttori e Interpreti dell'Università di Heidelberg* (IÜD, Universität Heidelberg 1996). Nel modello TETASTI, la categoria *Grammar* costituisce una macrocategoria autonoma, distinta da *Accuracy* e *Fluency*, diversamente da quanto avviene nella tassonomia SCATE e nelle *Annotation Guidelines* di Chiocchetti e De Camillis (2022), in cui *Grammar* è trattata come sottocategoria della *Fluency*. Tale scelta è motivata dalla volontà di evidenziare l'importanza degli aspetti grammaticali nella valutazione della traduzione di testi scientifici, operando una distinzione tra gli errori di *Grammar*, intesi come deviazioni grammaticali della lingua di arrivo, e gli errori di *Fluency*, che invece interessano la scorrevolezza della lingua di arrivo. In questo modo, TETASTI consente una valutazione più precisa e mirata delle dimensioni linguistiche e strutturali del testo tradotto. Pertanto, rispetto ai modelli presentati precedentemente, TETASTI presenta un adattamento originale: alcune categorie sono state aggiunte, altre modificate o affinate per rispondere alle

specifiche esigenze linguistiche, testuali e metodologiche del presente studio empirico (cfr. allegato 1).

Gli errori di *Accuracy* riguardano la corrispondenza informativa tra testo di partenza e testo di arrivo e vengono rilevati attraverso il confronto diretto tra i due testi. Gli errori di *Fluency* concernono la scorrevolezza e la naturalezza della lingua di arrivo, mentre quelli di *Grammar* ne riguardano la correttezza formale. Queste ultime due macrocategorie si concentrano quindi sulla qualità linguistica e strutturale del testo di arrivo: gli errori ad esse attribuibili emergono generalmente già analizzando solo il testo di arrivo; tuttavia, anche in questi casi il confronto con il testo di partenza rimane indispensabile per una valutazione completa. La comparazione, dunque, permette di verificare non solo la fedeltà informativa ma anche la correttezza formale del testo di arrivo, l'aderenza alle convenzioni del linguaggio scientifico e la corretta implementazione delle relative caratteristiche.

Gli errori di *Accuracy*, *Fluency* e *Grammar* possono essere annotati per uno stesso segmento di testo, poiché una categoria di errore può coesistere con un'altra. Tuttavia, si consiglia di annotare possibilmente un solo errore per segmento, soprattutto quando si tratta di categorie appartenenti a uno stesso sottogruppo di una macrocategoria, ad esempio più sottocategorie di A/MisTra (errori di traduzione) o di F/Lex (errori riguardanti la scelta lessicale di singole parole o costrutti). L'annotazione di più errori sullo stesso segmento è ammessa quando si tratta di fenomeni distinti e non di interpretazioni diverse dello stesso errore: per uno stesso segmento, ad esempio, non può essere annotato contemporaneamente un errore di corrispondenza semantica (A/MisTra/SemUn) e un errore di scelta lessicale inadeguata al genere o al dominio considerati (F/Lex/LexCho). In tali casi, va selezionata la categoria che meglio rappresenta il fenomeno osservato. Al contrario, è possibile annotare più errori quando essi interessano aspetti diversi del segmento, come nel caso di un errore di corrispondenza semantica (A/MisTra/SemUn) e un errore di trasferimento della formattazione (A/Mech/CharFor), oppure un errore di scorrevolezza che comporta una riorganizzazione della frase (F/Sty/Disfl) in cui uno degli elementi risulta anche interessato da un errore di traduzione semanticamente errata (A/MisTra/SemUn) o da un tempo verbale grammaticalmente scorretto (G/WF/VF). In generale, l'annotazione di più errori sullo stesso segmento deve essere limitata ai casi strettamente necessari, privilegiando sempre la semplicità e la coerenza dell'analisi.

Di seguito per ogni macrocategoria vengono descritte le sottocategorie che le compongono. Per ogni sottocategoria viene fornita una descrizione seguita da almeno un esempio tratto dal corpus di articoli scientifici tradotti con DeepL e analizzati in questo studio. In ogni esempio, il segmento in tedesco (indicato con *de*) è seguito dal corrispondente in italiano (*it*). Successivamente è proposta

la soluzione corretta (preceduta da →) o si propone una breve spiegazione. La parte della frase coinvolta dall'errore e dalla correzione è sottolineata. Per alcune categorie vengono aggiunte indicazioni utili per l'annotazione.

Le linee guida TETASTI restano aperte a ulteriori estensioni o adattamenti, qualora emergano esigenze descrittive non coperte dalle categorie di seguito descritte.

5.3.5.2.1 Errori di Accuracy

Gli errori di *Accuracy* sono descritti come errori di informazione, nello specifico: il testo di arrivo e il testo di partenza non contengono le stesse informazioni. Pertanto, spesso gli errori di *Accuracy* sono interpretazioni errate, contraddizioni, spostamenti di significato, aggiunte o omissioni. La Tabella 1 illustra le differenti sottocategorie di *Accuracy*, che saranno illustrate nel dettaglio nei paragrafi seguenti.

Errori di Accuracy												
Addition	Omission	Untranslated	Do-not-translate	Foreign language	MisTranslation	Multi-word-expressions	Terminology	Missed explicitation	Misordering	Mechanical	Source error	Other
					Semantically unrelated	Komposita	Semantically unrelated		Word order	Punctuation		
					Part-of-speech	Ellipsis	Do-not-translate		Misplaced word	Capitalization		
					Word sense disambiguation					Character formatting		
					Partial Gender							
					Wrong prep/conj							

Tabella 1: Errori di *Accuracy*

I. Accuracy

I.I Addition (A/Add)

Definizione: nella frase del testo di arrivo compare un segmento non presente nella frase del testo di partenza. Questo errore può riguardare parole, parti di parole e numeri e si può configurare in due modi: il segmento interessato può essere una ripetizione di un segmento già presente o una aggiunta di un segmento “nuovo”. Questa categoria non riguarda la punteggiatura.

Esempio:

de: Diese »kulturelle[n] Interferenz[en]« (Koreik 1998: 82) entstehen [...]

it: Queste “interferenze culturali” (Koreik 1998: 82) (Koreik 1998: 82) nascono [...]

→ Nella frase di arrivo si verifica una ripetizione di un segmento.

I.II Omission (A/Omis)

Definizione: il segmento del testo di partenza non è presente nel testo di arrivo. Questo errore può riguardare una o più parole, parti di parole e numeri. Questa categoria non riguarda la punteggiatura.

Esempio:

de: Bei b) und f) kommt das sogenannte “si passivante” [...] mit Verb jeweils im Singular und im Plural zum Einsatz.

it: In b) e f) il cosiddetto “si passivante” [...] con il verbo al singolare e al plurale.

→ In b) e f) viene impiegato il cosiddetto “si passivante [...] con il verbo rispettivamente al singolare e al plurale.

Nota per l’annotazione: annotare l’errore nel punto della frase di arrivo in cui si inserirebbe la parola o le parole omesse.

I.III Untranslated (A/Untra)

Definizione: la traduzione del segmento del testo di partenza era necessaria, ma non è stata eseguita. Questo errore può riguardare una o più parole.

Esempio:

de: Abschließend werden relevante Perspektiven skizziert, die erste Anregungen diesbezüglich enthalten.

it: Abschließend werden relevante Perspektiven skizziert, die erste Anregungen diesbezüglich enthalten.

→ L'intero segmento del testo di partenza non è stato tradotto anche se la traduzione era necessaria.

I.IV Do-not-translate (A/DNT)

Definizione: la traduzione del segmento del testo di partenza non era necessaria, ma è stata eseguita. Questo errore può riguardare una o più parole.

Esempio:

de: [...] in seinem Aufsatz "On linguistic aspects of translation" [...]

it: nel suo saggio "Sugli aspetti linguistici della traduzione" [...]

→ nel suo saggio "On linguistic aspects of translation" [...]

I.V Foreign language (A/ForLa)

Definizione: il segmento del testo di partenza è stato tradotto impiegando la lingua inglese o parole che non esistono nella lingua di arrivo. Questo errore può riguardare una o più parole.

Esempio:

de: [...] nicht 'gelesen' werden [...]

it: [...] non sono 'read' [...]

→ [...] non sono 'lette' [...]

I.VI Mistranslation

Definizione: il segmento del testo di arrivo è stato tradotto in modo errato.

I.VI.I Semantically unrelated (A/MisTra/SemUn)

Definizione: il significato di una o più parole del testo di arrivo non è correlato al significato della parola o delle parole del testo di partenza e non ha senso nel contesto.

Esempio:

de: [...], die machtvolle gesellschaftliche Ungleichheitsverhältnisse [...] sowie die eigene Rolle der Pädagog/innen bei deren Reproduktion und Verfestigung zum Gegenstand macht.

it: [...], che rende potenti le relazioni di disuguaglianza sociale [...], nonché il ruolo degli educatori stessi nella loro riproduzione e nel loro consolidamento della materia.

→ [...], che abbia come oggetto le potenti relazioni di disuguaglianza sociale [...], nonché il ruolo degli educatori stessi nella loro riproduzione e nel loro consolidamento.

I.VI.II Part-of-speech (A/MisTra/POS)

Definizione: il segmento del testo di arrivo presenta una categoria lessicale (parti del discorso, *Wortarten*) errata, semanticamente non corrispondente a quella del testo di partenza.

Esempio:

de: Formen fragiler Kulturalität

it: forme di fragilità culturale

→ forme di fragile culturalità

I.VI.III Word sense disambiguation (A/MisTra/WSD)

Definizione: il testo di arrivo si riferisce a un significato diverso e non adeguato al testo di partenza.

Esempio:

de: Trainingsszenario

it: scenario di allenamento

→ scenario di formazione

I.VI.IV Partial (A/MisTra/Part)

Definizione: questa categoria si riferisce a errori di numero e/o genere errato, o a tempi e/o modi verbali sbagliati (ad esempio, un significato corretto ma un tempo o un modo sbagliato, o un significato quasi identico ma una sfumatura diversa a causa di un verbo modale) che dunque alterano parzialmente il significato trasferito nel testo di arrivo rispetto al testo di partenza. Può essere impiegata anche per gli articoli (determinativi/ indeterminativi) e le preposizioni (semplici/articolate).

Esempio:

de: die beiden Übersetzerinnen

it: i due traduttori

→ le due traduttrici

È importante distinguere tra errori di A/MisTra/Part e quelli di G/WF/Agree (cfr. III.II.I). Gli errori di quest'ultima categoria riguardano la concordanza

morfosintattica all'interno del segmento di arrivo: ad esempio il soggetto e il verbo non concordano correttamente in numero e genere. L'errore di A/MisTra/Part, invece, riguarda la fedeltà informativa: la frase è grammaticalmente corretta, ma il significato trasferito non corrisponde al testo di partenza, ad esempio per un errore di genere o numero. Di seguito sono riportati due esempi per chiarire tale distinzione. Questi casi non sono tratti dal corpus preso in esame, ma sono stati creati appositamente per rendere più chiara la distinzione e permettere un confronto comparativo tra le categorie di errori.

Esempi:

1. **de:** die beiden Übersetzerinnen wurden interviewt

it: le due traduttrici sono stati intervistati

→ le due traduttrici sono state intervistate → errore di G/WF/Agree poiché manca la concordanza tra soggetto e verbo;

2. **de:** die beiden Übersetzerinnen wurden interviewt

it: i due traduttori sono stati intervistati

→ le due traduttrici sono state intervistate → errore di A/MisTra/Part poiché la frase è grammaticalmente corretta, ma il genere non corrisponde a quello riportato nel testo di partenza.

I.VI.V Gender (A/MisTra/Gend)

Definizione: il linguaggio inclusivo è impiegato o trasmesso non correttamente nel testo di arrivo.

Esempio:

de: [...] die Schülerinnen und Schüler [...]

it: [...] gli studenti [...]

→ le studentesse e gli studenti: nella traduzione dell'espressione "die Schülerinnen und Schüler", è stata impiegata solo la forma maschile "studente". La forma femminile è stata omessa.

I.VI.VI Wrong prep/conj (A/MisTra/WroPr)

Definizione: errori riguardanti l'uso di preposizioni, sintagmi preposizionali o congiunzioni che veicolano un significato differente da quello del testo di partenza.

Esempio:

de: [...] seit einigen Jahren [...]

it: [...] per diversi anni [...]

→ [...] da diversi anni [...]

I.VII Multi-word expressions

Definizione: errori riguardanti la resa traduttiva di una o più parole composte

I.VII.I Komposita (A/MWE/Komp)

Definizione: errori riguardanti la traduzione di aggettivi composti (*Komposita*). Spesso la traduzione risulta essere troppo letterale, in quanto riproduce i componenti della parola composta nello stesso ordine oppure la scompone.

Esempio:

de: eine philosophisch-ethische Reflexion

it: una riflessione filosofico-etica

→ una riflessione etico-filosofica

de: Die sprachlich-kulturellen Anpassungen

it: Gli adattamenti linguistici e culturali

→ Gli adattamenti linguistico-culturali

I.VII.II Ellipsis (A/MWE/Ell)

Definizione: errori riguardanti una serie di parole composte con elisione dell'elemento comune (*elliptische Kompositareihen*⁴⁷): nella traduzione, l'elemento condiviso viene omesso in riferimento ad alcuni composti e applicato solo ad altri, invece di riferirsi correttamente a tutti i composti presi in esame.

Esempio:

de: Flow- und Blockadeerlebnissen

it: flusso e delle esperienze di blocco

→ esperienze di flusso e di blocco: “Erlebnissen” si riferisce a entrambi i concetti (“Flow” e “Blockade”), tuttavia nella traduzione viene erroneamente associato esclusivamente a “Blockade” (esperienze di blocco), mentre “flusso” rimane scollegato.

⁴⁷ <https://grammis.ids-mannheim.de/fragen/3966> (ultima consultazione: 20/12/2025)

I.VIII Terminology

Definizione: il termine specialistico presente nel testo di partenza è trasferito in modo errato nel testo di arrivo.

I.VIII.I Semantically unrelated (A/Termi/SemUn)

Definizione: il significato del termine del testo di arrivo non è correlato al significato del termine del testo di partenza e pertanto non ha senso nel contesto.

Esempio:

de: eine morische Segmentierung

it: una segmentazione morfica

→ una segmentazione moraica

I.VIII.III Do-not-translate (A/Termi/DNT)

Definizione: la traduzione del termine del testo di partenza non era necessaria, ma è stata eseguita.

Esempio:

de: den Regeln der Leichten Sprache

it: le regole del linguaggio semplice

→ le regole della Leichte Sprache: la traduzione del termine non era necessaria in quanto in questo caso si fa specifico riferimento al linguaggio facile tedesco, cioè la *Leichte Sprache*.

I.IX Missed explicitation (A/MisExpl)

Definizione: Le informazioni fornite dalla frase di partenza non sono sufficienti a garantirne la comprensione da parte del lettore di arrivo. Tale comprensione richiederebbe aggiunte attraverso la compresenza della parola in italiano e quella in tedesco tra parentesi, o viceversa. In altri casi può essere necessario esplicitare informazioni che sono implicite per il destinatario per il testo di partenza, ma non per il destinatario del testo di arrivo.

Esempio:

de: Auf der Website des Wellnessverbandes [...]

it: Sul sito web dell'associazione per il benessere [...]

→ Sul sito web dell'associazione tedesca per il benessere [...]: è stata aggiunta la parola “tedesca” per consentire al lettore nella lingua di arrivo di comprendere che si parla di un’associazione tedesca e non italiana.

I.X Misordering

Definizione: errori in riferimento all'ordine di una o più delle parole.

I.X.I Word order (A/MisOrd/WO)

Definizione: l'ordine di una o più parole nel segmento del testo di arrivo veicola un significato diverso rispetto a quello trasmesso dall'ordine delle parole nel corrispondente segmento del testo di partenza.

Esempio:

de: [...], desto mehrwertiger und heterogener bzw. hybrider erscheinen kulturelle Akteursfelder.

it: [...], più appaiono gli ambiti di azione degli attori culturali polivalenti ed eterogenei o ibridi.

→ [...], più gli ambiti di azione degli attori culturali appaiono polivalenti ed eterogenei o ibridi.

Nota per l'annotazione: impiegare questa categoria esclusivamente per quei casi in cui l'ordine della parola influenza il senso della frase o del segmento di frase. Per i casi in cui si intende modificare l'ordine delle parole per ragioni stilistiche, impiegare la categoria F/Sty/Disfl (cfr. II.III.I).

I.X.II Misplaced word (A/MisOrd/MW)

Definizione: un elemento di un composto presente nella frase di arrivo è collocato in una posizione tale da riferirsi a una parola diversa rispetto a quanto indicato nel segmento di partenza. L'errore riguarda più parole. In alcuni casi può rendersi necessario non solo un diverso posizionamento dell'elemento del composto, ma anche un adeguamento morfologico (ad esempio in termini di genere e numero).

Esempio:

de: [...] einem individualziel- und entwicklungsorientierten Coaching [...]

it: [...] il coaching individuale orientato agli obiettivi e allo sviluppo [...]

→ [...] il coaching orientato agli obiettivi individuali e allo sviluppo [...]:
individuale non si riferisce a coaching ma a obiettivi. La parola è stata concordata con la nuova parola di riferimento.

de: Anpassungsschwierigkeiten oder Konflikte [...]

it: Le difficoltà o i conflitti di adattamento [...]

→ Le difficoltà di adattamento o i conflitti [...]: la parola adattamento si riferisce a difficoltà, non a conflitti.

I.XI Mechanical

Definizione: errori di trasferimento meccanico non legati al trasferimento di contenuti.

I.XI.I Capitalization (A/Mech/Capit)

Definizione: errori relativi al trasferimento delle regole relative all'uso delle lettere maiuscole dal testo di partenza al testo di arrivo.

Esempio:

de: Die Großschreibung des Adjektivs ‚Weiß‘ [...]

it: La maiuscola dell'aggettivo “bianco” [...]

→ Le lettere maiuscole non sono state trasferite dal testo di partenza al testo di arrivo.

I.XI.II Punctuation (A/Mech/Punct)

Definizione: errori relativi al trasferimento della punteggiatura dal testo di partenza al testo di arrivo, come omissioni, aggiunte o modifiche della punteggiatura.

Esempio:

de: [...] kein Spezialformat (mehr) ist

it: [...] non è più un formato specializzato

→ [...] non è (più) un formato specializzato

Nota per l'annotazione: nel caso in cui si verifichi un'omissione, annotare l'errore nel punto della frase di arrivo in cui si inserirebbe la punteggiatura omessa.

I.XI.III Character formatting (A/Mech/CharFor)

Definizione: errori riguardanti la formattazione del carattere di una o più parole, non trasferita in modo corretto nel testo di arrivo. Sono inclusi errori riguardanti lo stile di font, il colore, la grandezza.

Esempio:

de: einfach über ein ergänzendes Adjektiv hinzugenommen werden können

it: possano essere semplicemente aggiunti tramite un aggettivo supplementare

→ possano essere semplicemente aggiunti tramite un aggettivo supplementare

I.XII Source error (A/SouErr)

Definizione: errori presenti nel testo di partenza. Questa categoria comprende anche errori presenti nel testo di partenza elaborato da Trados Studio per la traduzione sulla sua piattaforma. Tali errori emergono confrontando il testo di partenza elaborato dal CAT tool con il testo originale.

Esempio:

de: [...] an die Crtical Whiteness Studies [...]

it: [...] fa seguito ai Crtical Whiteness Studies [...]

→ [...] fa seguito ai Critical Whiteness Studies [...]: la parola è scritta in modo scorretto già nel testo di partenza.

de: Welche Lernziele, Trainingstypen und bieten sich an [...]

it: Quali obiettivi di apprendimento e quali tipi di formazione sono adatti a [...]

→ La corretta frase di partenza è la seguente: Welche Lernziele, Trainingstypen und Übungen bieten sich an, [...]. A questa corrisponde la seguente frase di arrivo corretta: Quali obiettivi di apprendimento, tipi di formazione ed esercizi sono adatti a [...]. Si tratta di errore presente nel testo di partenza del file elaborato da Trados Studio, ma non presente nel file originale del testo.

I.XIII Other (A/Oth)

Definizione: la traduzione non è corretta, ma il problema non può essere ricondotto a nessuna delle sottocategorie precedenti.

Esempio:

de: Abstract (Deutsch)

it: Abstract (tedesco)

→ Abstract (italiano): la traduzione è fedele ma non è funzionale e quindi corretta, in quanto si tratta di una traduzione in italiano di un testo originariamente redatto in tedesco.

5.3.5.2.2 Errori di *Fluency*

Gli errori di *Fluency* possono essere definiti come errori riguardanti la scorrevolezza della lingua di arrivo. Una buona traduzione per essere tale deve dare al lettore la percezione che il testo sia stato scritto originariamente in italiano. Gli errori di tale macrocategoria sono già visibili nel testo di arrivo, tuttavia, richiedono in ogni caso un confronto con il testo di partenza per accertarne la

natura. Nella Tabella 2 è possibile individuare le differenti sottocategorie appartenenti alla categoria principale *Fluency*, di seguito descritte nel dettaglio.

Errori di Fluency				
<u>Lexicon</u>	<u>Scientific language phenomena</u>	<u>Style</u>	<u>Design</u>	<u>Coherence</u>
<u>Lexical choice</u>	<u>Nominalization</u>	<u>Disfluency</u>	<u>Character formatting</u>	<u>Co-reference</u>
<u>Fügung</u>	<u>Impersonal</u>	<u>Register</u>	<u>Layout</u>	<u>Inconsistent punctuation</u>
		<u>Idiomatcity</u>		<u>Citation</u>

Tabella 2: Errori di *Fluency*

II. Fluency

II.I Lexicon

Definizione: errori nell'uso del vocabolario in lingua tedesca o italiana.

II.I.I Lexical choice (F/Lex/LexCho)

Definizione: una o più parole appartengono al lessico della lingua italiana o tedesca, tuttavia, esistono espressioni fisse maggiormente adeguate all'ambito specialistico e/o al genere testuale preso in esame.

Esempio:

de: Für die Äußerungen in der Beschreibung der Bildergeschichte [...] reichte jedoch ein einmaliges Auftreten aus.

it: Tuttavia, per gli enunciati nella descrizione della storia illustrata [...], è stata sufficiente una sola apparizione.

→ Tuttavia, per gli enunciati nella descrizione della storia illustrata [...], è stata sufficiente una sola occorrenza.

II.I.II Fügung (F/Lex/Fueg)

Definizione: la frase di partenza presenta una *Fügung*, la cui resa nella frase di arrivo è grammaticalmente corretta ma è possibile individuare una formulazione più consolidata nella lingua di arrivo rispetto a quella proposta.

Esempio:

de: Unter dem “Diskurs” wird hier eine Reihe von Texten verstanden, [...]

it: Il “discorso” si riferisce qui a una serie di testi [...]

→ Con il termine “discorso” si fa riferimento a una serie di testi [...]

II.II Scientific language phenomena

Definizione: errori riguardanti fenomeni linguistici tipici del linguaggio scientifico.

II.II.I Nominalization (F/SLP/Nomin)

Definizione: la frase o una parte di frase è grammaticalmente corretta, tuttavia, è necessaria la nominalizzazione di un segmento della frase per una resa ideale del linguaggio scientifico.

Esempio:

de: Die Bedeutung der Reflexion der eingesetzten Methoden wird diskutiert [...]

it: Viene discussa l'importanza di riflettere sui metodi utilizzati [...]

→ Viene discussa l'importanza della riflessione sui metodi utilizzati [...]

II.II.II Impersonal (F/SLP/Imp)

Definizione: la frase o una parte di frase è grammaticalmente corretta, tuttavia, è necessario l'uso di verbi alla forma impersonale per una resa ideale del linguaggio scientifico.

Esempio:

de: Zur Beantwortung wird untersucht, ob [...]

it: Per rispondere a questa domanda, analizzeremo se [...]

→ Per rispondere a questa domanda, si analizzerà se [...]

II.III Style

Definizione: errori riguardanti le scelte stilistiche o il grado di formalità del testo.

II.III.I Disfluency (F/Sty/Disfl)

Definizione: la frase o una parte della frase è grammaticalmente corretta, tuttavia risulta difficile da leggere e/o spesso da comprendere. Questo errore può riguardare una singola parola o più parole e può comportare modifiche più consistenti delle parole o, più semplicemente, una modifica dell'ordine delle stesse. Gli interventi non

sono finalizzati alla correzione del contenuto semantico, bensì al raggiungimento di una maggiore scorrevolezza del segmento coinvolto, includendo le modifiche per evitare ripetizioni secondo il requisito di variazione.

Esempio:

de: [...] denn sie entspricht den obigen nach Jakobson angeführten Charakteristika der ersten Übersetzungskategorie.

It: [...] in quanto corrisponde alle caratteristiche della prima categoria di traduzione sopra elencate secondo Jakobson.

→ [...] in quanto corrisponde alle caratteristiche, sopra elencate, della prima categoria di traduzione individuata da Jakobson.

Si raccomanda l'impiego di questa categoria anche nei casi in cui il cambio di genere di una o più parole sia una conseguenza indiretta della correzione di un errore di A/MisTra/SemUn (mancata corrispondenza semantica, cfr. I.VI.I), A/MisTra/WSD (traduzione errata di parole polisemiche, cfr. I.VI.III) o di F/Lex/LexCho (scelta lessicale non adeguata al genere e all'ambito specialistico, cfr. II.I.I). In tali casi, la modifica del genere non è dovuta a una mancata concordanza morfosintattica (G/WF/Agree, cfr. III.II.I) né a un trasferimento incorretto del genere (A/MisTra/Part, cfr. I.VI.IV), bensì al fatto che la correzione di segmento semanticamente errato rispetto al testo di partenza (A/MisTra/SemUn o A/MisTra/WSD) o lessicalmente non adeguato al linguaggio scientifico, al genere testuale e al dominio trattati (F/Lex/LexCho), comporta la necessità di modificare il genere di uno o più elementi ad esso collegati, al fine di rendere la frase di arrivo corretta e scorrevole.

Esempio:

de: Die Äußerungen wurden [...] aufgezeichnet

it: le dichiarazioni sono state registrate [...]

→ gli enunciati (F/Lex/LexCho) sono stati registrati (F/Sty/Disfl) [...]

Nota per l'annotazione: nei casi in cui la modifica dell'ordine delle parole comporti un cambiamento del significato dell'enunciato e un'alterazione della struttura informativa, si raccomanda l'impiego della categoria A/MisOrd/WO (cfr. I.X.I).

II.III.II Register (F/Sty/Reg)

Definizione: errori riguardanti il registro impiegato (formale o informale), che non corrisponde al grado di formalità richiesto dal genere testuale.

Esempio:

de: [...] und sich in die Situation der anderen Kursteilnehmer hineinversetzen.

it: [...] e si mettono nei panni degli altri partecipanti al corso.

→ [...] e si immedesimano negli altri partecipanti al corso.

II.III.III Idiomaticity (F/Sty/Idiom)

Definizione: la frase di partenza presenta un'espressione idiomatica la cui resa nella lingua di arrivo si realizza con una frase o una parte di frase che è grammaticalmente corretta ma che non risulta altrettanto idiomatica e quindi scorretta e/o difficile da comprendere.

Esempio:

de: Die in der Sowjetzeit kreierte Übersetzungen von Böll, Dürrenmatt, Frisch, aber auch manche von Feuchtwanger und Remarque, sind zum Himmel schreiend schlecht [...]

it: Le traduzioni di Böll, Dürrenmatt, Frisch, ma anche alcune di Feuchtwanger e Remarque, realizzate in epoca sovietica, gridano al cielo [...]

→ Le traduzioni di Böll, Dürrenmatt, Frisch, ma anche alcune di Feuchtwanger e Remarque, realizzate in epoca sovietica, sono vergognosamente scadenti [...]

II.IV Design

Definizione: errori relativi alla progettazione grafica o alla presentazione di un prodotto di traduzione, compresi la formattazione e la marcatura di caratteri, la strutturazione del testo e il layout generale della pagina.

II.IV.I Character Formatting (F/De/CharFor)

Definizione: le convenzioni linguistiche del testo di arrivo richiedono una formattazione diversa del carattere.

Esempio:

de: [...] einheitliche Übersetzung für Moor bzw. Moorbad [...]

it: [...] una traduzione standardizzata di Moor o Moorbad [...]

→ [...] una traduzione standardizzata di Moor o Moorbad [...]

Nota per l'annotazione: non impiegare questa categoria di errori per i casi in cui nella frase di arrivo non è stata rispettata la formattazione della frase di partenza. Per quei casi usare A/Mech/CharFor (cfr. I.XI.III).

II.IV.II Layout (F/De/Lay)

Definizione: errori riguardanti la strutturazione del testo e dei paragrafi, comprese le interruzioni di riga non necessarie e gli elenchi puntati non indicati.

Esempio:

de: Anschließend wurden pro Diskussionsgruppe 3-4 dieser Falle ausgewählt, [...]

it: Successivamente, per ogni gruppo di discussione

Sono stati selezionati 3-4 di questi casi [...]

→ Nella frase di arrivo si verifica un'interruzione di riga che spezza la frase. In questo caso, l'errore di layout è accompagnato da un errore di G/Orth/Capit, riguardante l'uso improprio della lettera maiuscola. In altri casi, potrebbero sorgere errori differenti come A/MisTra/SemUn, se la frase risulta semanticamente errata; oppure errori di G/Orth/Punct, per i casi in cui l'interruzione di riga viene percepita come la fine di una frase e viene aggiunto un punto fermo nella prima riga.

II.V Coherence

Definizione: questa categoria è utilizzata per l'annotazione di errori di coerenza a livello testuale che riguardano parti ampie del testo o l'intero testo.

II.V.I Co-reference (F/Coh/CoRef)

Definizione: errori relativi alla mancata corrispondenza di entità in frasi diverse (ad esempio, un pronome femminile che si riferisce a una persona maschile menzionata in una frase precedente). Tali errori non compromettono necessariamente la correttezza grammaticale della singola frase, ma la rendono comunque incoerente a livello di paragrafo o di testo. In questa categoria rientrano anche le incoerenze nell'uso dei tempi verbali (ad esempio, l'impiego del futuro semplice in contrasto con il presente indicativo utilizzato nelle frasi precedenti).

Esempio:

de: Bei einer genaueren Untersuchung zeigen sich jedoch Unterschiede. Diese beziehen sich auf die zur Verfügung gestellten Audio- und sonstigen Materialien [...]

it: Tuttavia, un esame più attento rivela delle differenze. Questi riguardano il materiale audio e altri materiali forniti [...]

→ Tuttavia, un esame più attento rivela delle differenze. Queste riguardano il materiale audio e altri materiali forniti [...]

de: Dieser Stil wird nun auf seine kollektiven Wirkungen und Implikationen im spezifischen Kontext untersucht

it: Questo stile verrà ora analizzato per i suoi effetti collettivi e le sue implicazioni in un contesto specifico

→ Questo stile viene ora analizzato per i suoi effetti collettivi e le sue implicazioni in un contesto specifico: la frase è grammaticalmente corretta, tuttavia risulta incoerente a livello testuale perché in precedenza è stato impiegato il tempo verbale presente indicativo.

II.V.II Inconsistent punctuation (F/Coh/InPu)

Definizione: errori riguardanti l'uso incoerente di segni ortografici quali le virgolette e l'apostrofo.

Esempio:

it: L'obiettivo era invece [...]

→ L'obiettivo era invece [...]

it: "Lingue specialistiche"

→ "Lingue specialistiche": questo errore si verifica frequentemente nei testi tradotti automaticamente, poiché i sistemi di TA tendono a utilizzare le virgolette dritte (" ") invece di quelle tipografiche (“ ”). Tale discrepanza genera una incoerenza visiva e formale, soprattutto quando il *post-editor* interviene nel testo, rendendo evidente la disomogeneità tra parti generate automaticamente e parti modificate manualmente.

Nota per l'annotazione: se l'errore riguarda l'uso delle virgolette, annotare le virgoletta aperta e quella chiusa come un unico errore.

II.V.III Citation (F/Coh/Cit)

Definizione: errori riguardanti i riferimenti bibliografici, la cui forma è incoerente.

Esempio:

de: [...] der entsprechenden Herkunft“ (ebd.)

it: [...] dell'origine corrispondente” (ibidem)

→ [...] dell'origine corrispondente” (ibid.): la traduzione risulta incoerente in quanto per le precedenti occorrenze della parola “ebd.” è stata impiegata la formula abbreviata “ibid.” e non quella estesa “ibidem”.

5.3.5.2.3 Errori di *Grammar*

Gli errori di *Grammar* riguardano la correttezza grammaticale o il rispetto delle norme grammaticali della lingua di arrivo. Il segmento coinvolto da tali errori risulta problematico già alla prima lettura del testo di arrivo; una volta individuato l'errore, il segmento di arrivo viene confrontato con il segmento di partenza per accertarne la natura e la correttezza. Di seguito è presente una tabella, all'interno della quale è possibile individuare le differenti sottocategorie appartenenti alla categoria principale *Grammar*, che vengono di seguito descritte nel dettaglio.

Errori di Grammar	
Orthography	Word form
Spelling	Agreement
Capitalization	Verb form
Punctuation	
Whitespace	

Tabella 3: Errori di *Grammar*

III. Grammar

III.I Orthography

Definizione: errori relativi alle convenzioni di scrittura della lingua italiana o tedesca.

III.I.I Spelling (G/Orth/Spe)

Definizione: errori ortografici, ad esempio l'ortografia riguardante la combinazione di parole o segni diacritici (accenti per l'italiano e Umlaut per il tedesco).

Esempio:

de: Angaben zum Vorkommen von [...]

it: informazioni sulla occorrenza di [...]

→ informazioni sull'occorrenza di [...]

III.I.II Capitalization (G/Orth/Capit)

Definizione: errori relativi all'impiego della lettera maiuscola. Il suo trasferimento dal testo di partenza viene trattato separatamente. Gli errori ortografici sono visibili a livello della lingua di arrivo senza necessità di controllare la lingua di partenza.

Esempio:

de: Kurz gesagt: Was sich angenehm anfühlt [...]

it: In breve: Cìò che è piacevole [...]

→ In breve: cìò che è piacevole [...]

III.I.III Punctuation (G/Orth/Punct)

Definizione: Errori relativi ai segni di punteggiatura. Questa categoria di errori non si riferisce agli errori di punteggiatura dovuti a una trascrizione errata. Gli errori ortografici sono visibili a livello della lingua di arrivo, senza necessità di controllare la lingua di partenza.

Esempio:

de: nicht jedoch _ bis auf wenige Ausnahmen _ [...]

it: ma non _ con qualche eccezione _ [...]

→ ma non₂ con qualche eccezione₂ [...]: l'uso del trattino breve per gli incisi non è usuale secondo le regole grammaticali italiane.

III.I.IV White space (G/Orth/WS)

Definizione: errori relativi all'impiego della spaziatura. Questi errori sono visibili a livello della lingua di arrivo senza necessità di controllare la lingua di partenza.

Esempio:

it: Come in *Aspekte*, anche qui [...]

→ Come in *Aspekte*, anche qui [...]

III.II Word form

Definizione: la forma di una o più parole è sbagliata.

III.II.I Agreement (G/WF/Agree)

Definizione: errori riguardanti l'accordo tra due o più parole. Può riguardare l'accordo tra nome e aggettivo, nome e articolo e verbo e soggetto.

de: Allerdings muss sich dann vergegenwärtigt werden, dass die Forderung, "einen gut lesbaren Text" (Maaß 2015: 130) zu erschaffen, dann metaphorisch aufzuschlüsseln ist, da [...]

it: Tuttavia, occorre tenere presente che il requisito di produrre "un testo facilmente leggibile" (Maaß 2015: 130) può essere scomposta metaforicamente, poiché [...]

→ Tuttavia, occorre tenere presente che il requisito di produrre “un testo facilmente leggibile” (Maaß 2015: 130) può essere scomposto metaforicamente, poiché [...]

Nota per l’annotazione: se l’errore di accordo si verifica come conseguenza di una correzione di un errore di corrispondenza semantica (A/MisTra/SemUn, cfr. I.VI.I), di una traduzione errata di parola polisemica (A/MisTra/WSD, cfr. I.VI.III) o di una scelta lessicale non adeguata al genere e all’ambito specialistico (F/Lex/LexCho, cfr. II.I.I), non usare questa categoria per annotare l’errore di accordo, bensì impiegare la categoria F/Sty/Disfl (errore di scorrevolezza, cfr. II.III.I).

III.II.II Verb form (G/WF/VF)

Definizione: errori riguardanti l’uso grammaticalmente scorretto di tempi o forme verbali.

de: dieser Umstand führt uns zur Schlussfolgerung, dass das Hauptobjekt der Übersetzung der Diskurs ist.

it: questo porta a concludere che l’oggetto principale della traduzione è il discorso.

→ questo porta a concludere che l’oggetto principale della traduzione sia il discorso.

Nota per l’annotazione: se si riscontra un errore di trasferimento del tempo verbale (ad esempio, l’uso del presente al posto dell’imperfetto richiesto dal testo di partenza), utilizzare la categoria A/MisTra/Part (cfr. I.VI.IV).

Si riportano di seguito alcune regole generali di annotazione, elaborate per la presente ricerca sulla base dell’esperienza di annotazione e dell’analisi delle categorie descritte nell’elaborato. Queste indicazioni mirano a garantire uniformità, precisione e coerenza nella fase di annotazione, indipendentemente dallo strumento impiegato (file Word o piattaforme dedicate come *WebAnno* o *INCEpTION*).

Regole generali di annotazione

1. Selezionare sempre una sottocategoria, non una categoria principale.
2. Si raccomanda di limitare di casi di annotazione multipla per un segmento ai casi strettamente necessari, soprattutto quando i problemi individuati appartengono a uno stesso sottogruppo (ad esempio, più sottocategorie di A/MisTra, traduzione errata, o di F/Lex, errori riguardanti le scelte lessicali). È tuttavia possibile annotare più errori sullo stesso segmento quando si tratta di fenomeni distinti e indipendenti, e non di diverse interpretazioni dello stesso errore. Ad esempio, un segmento può essere annotato

contemporaneamente come errore di traduzione semanticamente errata (A/MisTra/SemUn) e errore di trasferimento della formattazione (A/Mech/CharFor).

3. Se un segmento contiene più di un problema, evidenziarlo tante volte quanti sono i problemi, una volta per ogni problema.
4. Se si individua un errore che non può essere ricondotto a nessuna delle categorie già presenti, si raccomanda l'uso delle categorie *Other*: nel caso di errori di *Accuracy* usare la categoria A/Oth già esistente; per gli errori di *Grammar* e *Fluency*, è possibile creare le categorie G/Oth e F/Oth.
5. In caso di dubbio in riferimento all'annotazione, inserire nel commento un doppio punto interrogativo “??” per indicare questa incertezza.
6. Se lo stesso problema si verifica più di una volta, selezionare ogni occorrenza separatamente.
7. Se l'annotazione con l'indicazione della categoria di errore non è sufficiente a comprendere adeguatamente la problematica riscontrata, aggiungere un commento all'annotazione, impiegando anche fonti esterne come link.
8. Se si esegue l'annotazione degli errori sul file Word dell'output della TA, annotare gli errori nella sezione commenti denominandoli mediante un sistema codificato: la categoria principale è indicata con G (*Grammar*), A (*Accuracy*) o F (*Fluency*), seguita da una barra (/) e dall'acronimo che accompagna il nome delle categorie delle linee guida. Ad esempio, A/MisTra/WroPr segnala un errore nell'uso delle preposizioni o delle congiunzioni.
9. Se si esegue l'annotazione degli errori su un tool di annotazione come WebAnno o INCEpTION, annotare gli errori di *Accuracy* nel segmento del testo di arrivo e nel corrispettivo segmento di partenza; annotare, invece, gli errori di *Fluency* e di *Grammar* esclusivamente nel testo di arrivo.

5.3.6 Post-editing

La fase di *post-editing* è stata condotta su Trados Studio, procedendo con la correzione di un testo per volta. Ogni file conteneva l'output della TA, generato da DeepL Pro direttamente nell'editor del CAT tool. Una volta aperto il file, il testo risultava già segmentato secondo i criteri di segmentazione propri di Trados Studio e la correzione è stata effettuata procedendo segmento per segmento.

Come menzionato precedentemente, a ciascun progetto era stata associata una memoria di traduzione: di conseguenza, ogni volta che un segmento veniva confermato, questo si aggiungeva automaticamente alla memoria, consentendo l'eventuale recupero di possibili corrispondenze.

Parallelamente, a ciascun progetto era stato collegato anche un database terminologico, arricchito progressivamente durante la fase di *post-editing*: i termini individuati e tradotti sono stati inseriti man mano, così da garantire coerenza terminologica all'interno dei testi, oltre a facilitare il processo di traduzione.

Per ottimizzare la gestione del processo traduttivo, il lavoro è stato svolto su due schermi: su un monitor era visualizzato il file Word contenente l'output annotato con gli errori rilevati nella fase precedente, mentre sull'altro schermo si procedeva con la correzione su Trados Studio. In tal modo è stato possibile confrontare costantemente le annotazioni, verificarle ed eventualmente modificarle, integrarle o eliminarle. Inoltre, durante la fase di *post-editing* sono stati consultati diversi strumenti esterni che hanno svolto un ruolo determinante nelle scelte traduttive e terminologiche e nel miglioramento della resa:

- Dizionari online bilingue: Dizionario di Tedesco – il Sansoni Tedesco⁴⁸;
- Dizionari online monoligue: Duden Online⁴⁹ per il tedesco; Treccani⁵⁰ per l'italiano;
- Banche dati di tesi paralleli: Linguee⁵¹, Reverso Context⁵²;
- Motori di ricerca per la scrittura scientifica: Google Scholar⁵³;
- Strumenti per l'analisi di corpora: Sketch Engine⁵⁴;

I dizionari online bilingui, come il Sansoni Tedesco, sono stati impiegati per verificare corrispondenze lessicali tra le due lingue, mentre i dizionari monolingui, nello specifico Duden Online per il tedesco e Treccani per l'italiano, sono stati utilizzati per chiarire sfumature semantiche e assicurare un uso corretto e stilisticamente appropriato delle parole. Accanto a queste risorse, piattaforme come Linguee e Reverso Context hanno fornito esempi di traduzioni contestualizzate, utili per osservare il funzionamento delle espressioni in testi reali. Per quanto riguarda la specificità del linguaggio scientifico, è stato impiegato Sketch Engine, che ha permesso di analizzare il corpus in lingua italiana appositamente creato per questo studio e di estrarre collocazioni e strutture lessicali tipiche della scrittura scientifica, facilitando in questo modo le scelte traduttive. Come supporto integrativo all'uso del corpus, si è fatto ricorso a Google Scholar, che ha rappresentato uno strumento prezioso per verificare la frequenza d'uso e le combinazioni tipiche di termini ed espressioni nel discorso accademico. Infine, grazie a questo insieme di risorse, integrate e usate in modo complementare, il *post-editing* è stato condotto con un approccio sistematico, garantendo non

⁴⁸ https://dizionari.corriere.it/dizionario_tedesco/ (ultima consultazione: 20/12/2025)

⁴⁹ <https://www.duden.de/woerterbuch> (ultima consultazione: 20/12/2025)

⁵⁰ <https://www.treccani.it/vocabolario/> (ultima consultazione: 20/12/2025)

⁵¹ <https://www.linguee.it/> (ultima consultazione: 20/12/2025)

⁵² <https://context.reverso.net/traduzione/> (ultima consultazione: 20/12/2025)

⁵³ <https://scholar.google.com/> (ultima consultazione: 20/12/2025)

⁵⁴ <https://www.sketchengine.eu/> (ultima consultazione: 20/12/2025)

solo la correzione degli errori individuati, ma anche il raggiungimento di una maggiore precisione terminologica e aderenza agli standard stilistici del linguaggio scientifico.

Nonostante gli interventi di *pre-editing*, nella fase di *post-editing* sono stati riscontrati errori riconducibili alla segmentazione del testo di partenza operata da Trados studio. Ad esempio, nell'output generato dal sistema di TA si è constatata la frequente aggiunta di punteggiatura superflua nei riferimenti bibliografici. Un esempio di questo tipo di errore è presente nella Figura 24, relativa a un segmento estrapolato da un articolo della rivista *Informationen Deutsch als Fremdsprache*.

target di studenti. La memoria semantica è organizzata attraverso la categorizzazione, cioè le nuove informazioni vengono memorizzate assegnandole a un termine generico (cfr. Edelmann 2000): 125). Se questo viene

Figura 24: Esempio di punteggiatura superflua nel file Word

Il CAT tool suddivide la frase in corrispondenza dei due punti (:) interpretandoli come l'inizio di un nuovo segmento. Tale operazione costituisce un errore di segmentazione, poiché si tratta di un riferimento bibliografico in cui i due punti non indicano l'inizio di un nuovo segmento, ma specificano la pagina della pubblicazione da cui è stata estrapolata la citazione.

Nella Figura 25 è riportata la schermata dell'editor di testo di Trados Studio con il segmento del testo di partenza sulla sinistra e il rispettivo segmento del testo di arrivo sulla destra, già illustrato precedentemente nella Figura 24. La Figura 25 evidenzia l'errata segmentazione, con il segmento spezzato in corrispondenza dei due punti, come descritto in precedenza. Nel testo di arrivo viene inoltre inserita una parentesi tonda di chiusura perché il sistema di TA suppone la fine del segmento sulla base della segmentazione. Per questo motivo, nel testo di arrivo è possibile constatare l'aggiunta di punteggiatura superflua.

Das semantische Gedächtnis wird über Kategorisierung organisiert, d.h. die Speicherung von neuen Informationen erfolgt über die Zuordnung zu einem Oberbegriff (vgl. Edelmann 2000: 125).	NMT	La memoria semantica è organizzata attraverso la categorizzazione, cioè le nuove informazioni vengono memorizzate assegnandole a un termine generico (cfr. Edelmann 2000): 125).

Figura 25: Esempio di punteggiatura aggiunta nell'editor di testo

In fase di *post-editing* si è intervenuto quindi unendo tutti i segmenti spezzati erroneamente a livello dei riferimenti bibliografici, correggendo la segmentazione errata e rimuovendo l'errore di punteggiatura superflua. Tale operazione è stata resa possibile grazie alla funzione "unisci segmenti", accessibile tramite clic con il tasto destro del mouse dopo aver selezionato il primo segmento da unire e, tenendo premuto il tasto *Ctrl*, il secondo segmento (cfr. Fig. 26). Nel file Word, questo errore è stato annotato come A/Mech/Punct.

17	Wie die jüngste Datenerhebung des Netzwerks Deutsch nahelegt, ist zwischen 2010 und 2015 mit weiterhin anhaltender Tendenz in zahlreichen Ländern ein Anstieg der Deutschlernerzahlen zu verzeichnen.		NMT	Come suggerisce la più recente raccolta di dati del Netzwerk Deutsch , il numero di studenti di tedesco ha continuato ad aumentare in molti Paesi tra il 2010 e il 2015.
18	Bei einer Gesamtzahl von weltweit ca. 15,4 Millionen Lernern (Stand:		NMT	Con un numero totale di circa 15,4 milioni di studenti in tutto il mondo (come da:)
1	Copia testo di origine nei segmenti di destinazione	CTRL+INS	NMT	2015), tuttavia si possono osservare differenze regionali.
2	Cancella segmento di destinazione	ALT+CANC	NMT	Mentre l'interesse per il tedesco è aumentato in molti Paesi dell'America Latina, dell'Asia e dell'Africa, nonché in Medio Oriente, ad esempio, alcuni Stati della CSI stanno registrando un calo, in alcuni casi significativo (cfr. Ufficio federale degli affari esteri 2015):
2	Unisci segmenti	CTRL+ALT+S	NMT	6€).
2	Modifica stato segmento		NMT	L'ultima indagine Eurydice conferma, almeno in parte, il crescente interesse per l'Europa.
2	Blocca segmenti	CTRL+L	NMT	

Figura 26: Funzione *Unisci segmenti* in Trados Studio

Inoltre, in alcuni casi è stato necessario intervenire sulla segmentazione del testo di partenza con la funziona “dividi segmento” in quanto alcuni segmenti presentavano più frasi all’interno. Sebbene questo tipo di errore non compromettesse la qualità della traduzione, si è scelto di dividere i segmenti per rendere più agevole la fase di *post-editing*.

Nel capitolo seguente saranno presentati e discussi i risultati emersi dalla fase di annotazione degli errori e dall’attività di *post-editing*.

Studio empirico: analisi dei risultati e discussione

Il presente capitolo riporta l'analisi dei risultati emersi dal *post-editing* dei testi sottoposti alla TA e dall'annotazione degli errori individuati. Il capitolo, introdotto dalla descrizione della procedura adottata per l'analisi degli errori, è strutturato in tre sezioni principali, ciascuna dedicata a uno dei progetti di traduzione sviluppati nello studio. All'interno di ogni sezione, viene presentata un'analisi complessiva dei testi che compongono il progetto di traduzione. In particolare, i dati relativi ai singoli testi, disponibili negli allegati 2-4, vengono combinati per generare statistiche complessive, consentendo così di calcolare medie, distribuzioni e tendenze degli errori in maniera comparativa all'interno di ogni progetto. Completano l'analisi i dati relativi all'*effort* temporale e a quello tecnico (cfr. 3.6).

6.1 Procedura adottata per l'analisi degli errori

Inizialmente viene effettuata un'analisi quantitativa degli errori individuati sulla base della tassonomia TETASTI (cfr. 5.3.5); ciò consente di ottenere una panoramica dei pattern di errore più e meno frequenti a livello globale. Per ogni progetto l'analisi viene riassunta in una tabella composta da due colonne: la colonna di sinistra riporta le singole categorie di errore; la colonna di destra il numero di annotazioni per ciascuna categoria in ordine decrescente di frequenza (cfr. ad es. Tab. 4 – 6.2.2). Le righe della tabella sono evidenziate con tre diversi colori, al fine di facilitare il riconoscimento delle categorie principali: il verde per la categoria *Fluency*, il giallo per gli errori di *Accuracy* e il rosso per la categoria *Grammar*. Alla tabella segue un commento sugli errori annotati, corredato da alcuni esempi particolarmente significativi tratti dai testi sottoposti a *post-editing*. Al fine di mostrare concretamente la tipologia di errore, per ciascun esempio sono riportati il segmento dal testo originale (TO), la proposta del sistema di traduzione automatica (TA) e la versione rielaborata in fase di *post-editing* (PE). Le parti interessate dall'errore sono state sottolineate.

In seguito, vengono riportati i dati riguardanti lo sforzo temporale e quello tecnico relativi a ogni progetto. L'indicazione temporale del totale delle sessioni di *post-editing* è stata calcolata mediante l'impiego del plug-in di Trados Studio Quality. Questo dato permette di valutare il *temporal effort*, ovvero il tempo impiegato per lo svolgimento del *post-editing*. Infine, vengono presentati i dati forniti dal report di Post-Edit Compare, un ulteriore strumento di Trados Studio qui impiegato. Prendendo in esame tutti i testi del progetto simultaneamente, l'analisi dei report di Post-Edit Compare permette di valutare il *technical effort* complessivo, inteso come la quantità di interventi manuali necessari per portare l'output della TA al livello qualitativo desiderato, offrendo

una prospettiva dell'efficacia e dell'impatto del *post-editing* sull'intero progetto di traduzione analizzato. In questi report è possibile confrontare tre versioni di ciascun segmento: l'output originario prodotto dalla macchina, la versione post-editata e una rappresentazione comparativa che evidenzia visivamente le differenze tra le prime due (cfr. Fig. 27).

Target (Original)	Target (Updated)	Target (Comparison)	PEMp
_Schindler_Literarisch.20210517.docx (filtered: 424)			
Kirsten Schindler	Kirsten Schindler		100% Edit-Dist.: 0 Max chars: 17
<cf> Letteratura </cf> <cf> Traduzione </cf>	<cf> Traduzione letteraria </cf>	<cf> Letteratura </cf> <cf> Traduzione letteraria </cf>	15.38% Edit-Dist.: 22 Max chars: 26
<cf> A </cf> <cf> speciale </cf> <cf> forma </cf> <cf> di </cf> <cf> scrittura </cf>	<cf> Una </cf> <cf> forma <cf> speciale </cf> <cf> di </cf> <cf> scrittura </cf>	<cf> AUna </cf> <cf> speciale </cf> forma <cf> formaspeciale </cf> <cf> di </cf> <cf> scrittura </cf>	58.54% Edit-Dist.: 17 Max chars: 41
<cf> 1 </cf> <cf> Introduzione </cf>	<cf> 1 </cf> <cf> Introduzione </cf>		100% Edit-Dist.: 0 Max chars: 20

Figura 27: Report di Post-Edit Compare: analisi dei segmenti – testo n. 6 dal progetto n. 2

In tal modo è possibile osservare non soltanto quali porzioni del testo siano state modificate, ma anche l'entità e la natura degli interventi effettuati⁵⁵.

L'analisi è supportata da indicatori quantitativi che mirano a oggettivare il grado di modifica. Il parametro più rilevante, riportato nella Figura 27 nella colonna a destra, è il PEMP (*Post-Edit Modifications percent*), che esprime la percentuale di parole del testo rimaste invariate a seguito del *post-editing*: un valore pari al cento per cento segnala l'accettazione integrale della proposta della macchina, mentre percentuali inferiori denotano un incremento progressivo dell'intervento umano.

A tale misura si affianca la cosiddetta *edit distance*, che calcola il numero di caratteri modificati in rapporto alla lunghezza complessiva del segmento. Quest'ultimo parametro viene contestualizzato dal valore denominato *Max chars*, che indica la lunghezza massima del segmento in caratteri.

Oltre all'analisi dei segmenti, il report include ulteriori sezioni. La Figura 28, estrapolata da un report di Post-Edit Compare, presenta, nella parte sinistra, la sezione *Translation Modifications*, che riporta il numero totale di segmenti e parole analizzati e la quantità di quelli effettivamente modificati, ossia rivisti o corretti dal *post-editor*. Per ciascuna di queste due categorie, si individuano le quantità di segmenti tradotti automaticamente (*Automated Translation*) e la quantità di segmenti per i quali la macchina non ha proposto una traduzione, e che quindi sono stati inseriti ex novo (*New*). In questo caso 6.220 segmenti sono stati generati automaticamente, mentre solo per 32 segmenti non si sono registrate proposte dal sistema di TA (*New*). Vengono inoltre indicati i

⁵⁵ Per questioni di spazio, in questa sede non saranno illustrate e commentate le modifiche apportate a livello dei singoli segmenti di tutti i testi presi in esame.

segmenti e le parole interessati da *perfect match*, *context match*, ripetizioni e *fuzzy match* suddivisi nelle seguenti fasce: 100%, 95-99%, 85-94%, 75-84% e 50-74%. In questo caso i valori sono pari a zero e indicano che non si registrano contributi da parte della memoria di traduzione. Il grafico a colonne sottostante rappresenta visivamente i dati riguardanti i segmenti totali e quelli modificati: la colonna rossa indica i segmenti modificati; quella verde scuro il totale dei segmenti. Il grafico considera esclusivamente i segmenti tradotti automaticamente e registra come modificati anche i segmenti accettati senza correzioni. Per questo motivo, il numero dei segmenti modificati coincide con il totale dei segmenti tradotti automaticamente, includendo anche segmenti che sono stati confermati così come proposti dal sistema di TA. I segmenti *New* non sono inclusi nel grafico.

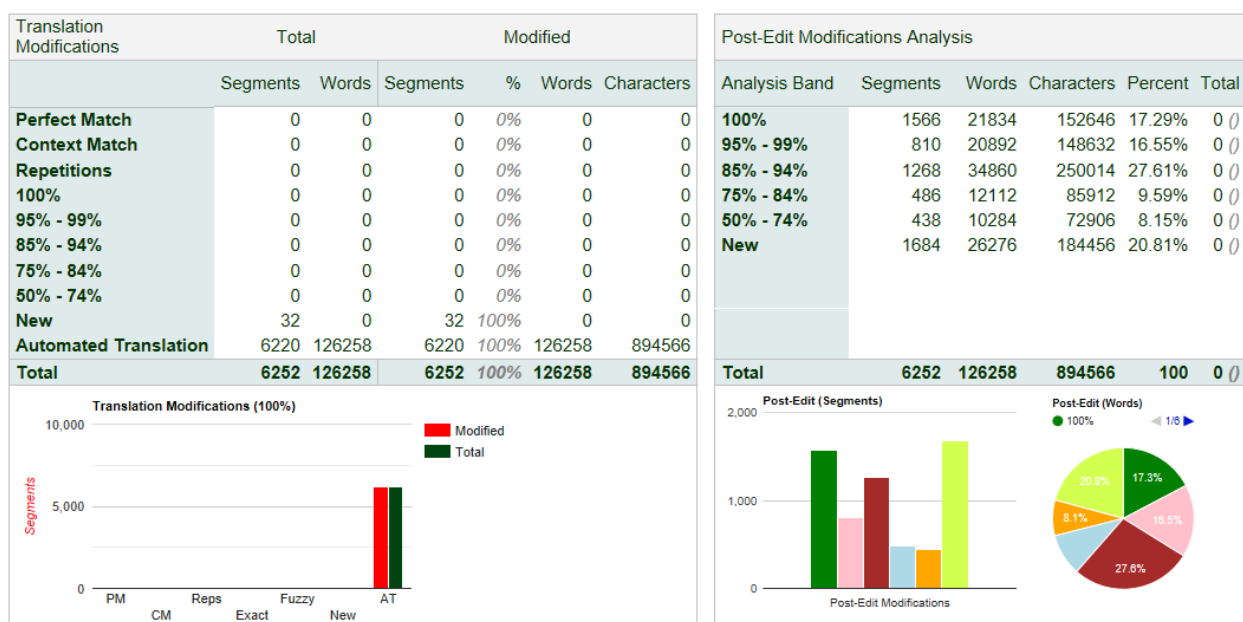


Figura 28: Report di Post-Edit Compare: *Translation Modifications* e *Post-Edit Modification Analysis* – progetto n. 3

A destra, la Figura 28 riporta anche la sezione del report in cui le percentuali di PEMp, come già accennato sopra, sono suddivise in sei fasce: 100%, 95-99%, 85-94%, 75-84%, 50-74% e *New*, dove 100% sono i segmenti che in fase di *post-editing* sono rimasti del tutto invariati. La fascia 95-99% include i segmenti che hanno richiesto interventi minimi, come modifiche lessicali mirate. Nella fascia 85-94% rientrano i segmenti che hanno necessitato di modifiche di entità lieve, pur mantenendo una struttura complessivamente corretta. La fascia 75-84% si riferisce a porzioni di testo che hanno richiesto correzione di media portata. La fascia 50-74% comprende segmenti per i quali la TA ha fornito solo un contributo parzialmente accettabile, risultando utile come base di partenza ma richiedendo un grado medio-alto di riformulazione. Infine, la categoria *New* raccoglie i segmenti con un livello di correzione compreso tra il 49% e lo 0%, cioè fino a rielaborazioni totali, evidenziando un output non adeguato. La fascia *New* di PEMp si distingue dalla categoria *New*

indicata nella sezione *Translation Modifications* in quanto quest'ultima indica i segmenti non tradotti automaticamente. Per ogni fascia di PEMp sono indicate le quantità di segmenti, parole e caratteri coinvolti. Sempre nella Figura 28, vengono rappresentati questi dati in forma grafica mediante un diagramma a torta e uno a colonne. Il diagramma a torta mostra la distribuzione percentuale di parole e caratteri, mentre il diagramma a colonne rappresenta la distribuzione dei segmenti. I colori agevolano l'individuazione delle fasce: verde scuro per il 100%, rosa per la fascia 95-99%, bordeaux per la fascia 85-94%, azzurro chiaro per la fascia 75-84%, arancione per la fascia 50-74% e verde chiaro per la categoria *New*.

Le indicazioni fornite dall'analisi complessiva consentono di delineare un quadro sufficientemente chiaro e articolato del comportamento del sistema di TA e del successivo intervento umano. I dati raccolti offrono una base concreta per interpretare le principali tendenze emerse e per valutare, in termini sia quantitativi sia qualitativi, l'efficacia del processo di *post-editing* nei diversi progetti analizzati.

6.2 Progetto n. 1: *Informationen Deutsch als Fremdsprache*

Di seguito (Tabella 4) viene riportata l'analisi complessiva dei testi che compongono il progetto di traduzione *Informationen Deutsch als Fremdsprache*. Il numero di *tokens* complessivo del progetto ammonta a 64.224.

6.2.1 Analisi degli errori⁵⁶

Categoria di errore	Somma delle annotazioni
F/Coh/InPu	1516
F/Sty/Disfl	807
A/MisTra/SemUn	794
F/Lex/LexCho	777
A/Termi/SemUn	291
A/Untra	213
A/DNT	198
A/Omis	192
G/Orth/Capit	173
A/ForLa	162
A/MisTra/WSD	138
A/MisTra/Part	117
G/Orth/Punct	111
F/SLP/Nomin	106
F/Sty/Reg	86
A/SouEr	84

⁵⁶ Le analisi dei singoli testi sono state riportate nell'allegato 2.

G/Orth/WhiSpa	81
A/MisOrd/WO	64
A/Add	64
A/MisTra/WroPr	54
A/MisExpl	48
A/MisTra/POS	45
G/WF/Agree	45
F/Lex/Fue	45
F/Coh/Cit	41
G/Orth/Spel	39
A/Mech/Punct	32
A/Termi/DNT	29
F/SLP/ForImp	28
A/MisTra/Gend	28
A/MisOrd/MW	25
F/Coh/CoRef	21
F/De/CharFor	16
F/De/Layo	11
A/Mech/CharFor	9
A/MWE/Ell	8
A/MWE/Komp	8
G/WF/VF	6
F/Sty/Idiom	6
A/Mech/Capit	4
A/Oth	0

Tabella 4: Analisi quantitativa degli errori per progetto n. 1

Complessivamente si osservano 3.460 errori di *Fluency*, 2.607 errori di *Accuracy* e 455 errori di *Grammar*. Il totale per l'intero progetto di traduzione ammonta a 6.522 errori. La categoria più rappresentata è la *Fluency*, principalmente a causa dell'elevata frequenza degli errori di F/Coh/InPu, che ammontano a 1.516 occorrenze. Questa categoria concerne l'uso incoerente di segni ortografici, quali l'apostrofo e le virgolette (cfr. esempio n. 1).

Esempio n. 1 (Prog1_Testo1 – F/Coh/InPu)

TA: l'utente può aver imparato

PE: l'utente può aver imparato [...]

Questo tipo di errore si riscontra frequentemente nei testi tradotti automaticamente, poiché i sistemi di TA tendono a utilizzare le virgolette dritte (" ") invece di quelle tipografiche (“ ”). Tale discrepanza genera un'incoerenza visiva e formale, che diventa evidente quando il *post-editor* interviene su singoli segmenti del testo per correggere altri tipi di errori e, nel farlo, utilizza virgolette o apostrofi tipografici, mettendo così in risalto la disomogeneità rispetto ai segmenti non

modificati, che conservano le virgolette dritte prodotte dal sistema di TA. Dalle analisi quantitative dei dati contenuti nell'allegato 2 emerge che F/Coh/InPu è la tipologia di errore più frequente in tutti i testi. Tuttavia, tali errori non incidono direttamente sulla qualità della traduzione del linguaggio scientifico, trattandosi di problematiche non riguardanti i contenuti e potenzialmente riscontrabili trasversalmente in diversi domini, generi testuali e coppie linguistiche trattati della TA.

Seguendo l'ordine decrescente di occorrenze, si riscontrano 807 errori di F/Sty/Disfl e 794 errori di A/MisTra/SemUn. Si tratta di errori riguardanti, nel primo caso, la scorrevolezza e la resa dello stile del linguaggio scientifico quotidiano nella lingua di arrivo (cfr. esempio n. 2) e, nel secondo caso, la mancata corrispondenza semantica tra testo di partenza e testo di arrivo (cfr. esempio n. 3). Queste sono tra le categorie più rappresentate nei testi post-editati, collocandosi tra il secondo e il quinto posto secondo le tabelle quantitative riportate nell'allegato 2.

Esempio n. 2 (Prog1_Testo3 – F/Sty/Disfl)

TO: äußerten, dass es ihnen nicht leichtfiel, [...]

TA: hanno dichiarato di non trovare facile [...]

PE: hanno dichiarato di aver riscontrato delle difficoltà [...]

L'errore di F/Sty/Disfl riguarda la resa stilistica in italiano: la traduzione automatica è comprensibile, tuttavia, risulta meno naturale rispetto al registro tipico del linguaggio scientifico quotidiano. In fase di *post-editing* è stata introdotta una formulazione più scorrevole e adeguata allo stile richiesto.

Esempio n. 3 (Prog1_Testo2 – A/MisTra/SemUn)

TO: Vielmehr ist das Ziel, einen Einblick in die sprachlichen Fähigkeiten [...] zu erhalten.

TA: L'obiettivo è piuttosto comprendere le capacità linguistiche [...].

PE: L'obiettivo è piuttosto acquisire una visione delle capacità linguistiche [...].

Nell'esempio n. 3, l'errore di A/MisTra/SemUn consiste nella mancata corrispondenza semantica: la traduzione automatica rende “Einblick erhalten” con “comprendere”, riducendo la complessità del significato originale. Secondo il Duden⁵⁷ il termine indica la possibilità, concessa a un osservatore esterno, di accedere a un contenuto e di esaminarlo dall'interno, sia in forma di sguardo introduttivo sia come lettura o analisi mirata. La proposta di *post-editing* (“acquisire una visione delle capacità linguistiche”), pur non riproducendo integralmente questo tratto semantico, rappresenta comunque una soluzione più adeguata: reintroduce almeno in parte l'idea del processo

⁵⁷ <https://www.duden.de/rechtschreibung/Einblick> (ultima consultazione: 20/12/2025)

cognitivo insita nell'espressione originale e risulta più conforme al registro del linguaggio scientifico rispetto alla resa generica proposta dal sistema di TA.

Di particolare rilievo, invece, è la categoria F/Lex/LexCho (777 occorrenze), che riguarda l'impiego di espressioni idonee all'ambito specialistico e al genere testuale specifico, ossia l'articolo scientifico. Questa rientra tra le categorie che registrano più risultati collocandosi tra il secondo e il quarto posto nelle tabelle quantitative degli errori per i testi di questo progetto di traduzione, illustrate nell'allegato 2. L'esempio n. 4 presenta un caso di errore per questa categoria.

Esempio n.4 (Prog1_Testo3 – F/Lex/LexCho)

TO: Präsenzlernen

TA: l'apprendimento faccia a faccia

PE: l'apprendimento in presenza

La terminologia rappresenta un elemento centrale del linguaggio scientifico: a questo proposito, nel presente progetto si identificano 291 errori di A/Termi/SemUn, relativi alla traduzione errata di termini (cfr. esempio n. 5), e 29 errori di A/Termi/DNT, categoria impiegata per indicare termini che non necessitano di traduzione (cfr. esempio n. 6). Come emerge dalle tabelle quantitative dei singoli testi, i valori più alti per la categoria A/Termi/SemUn si riscontrano nei testi n. 1, n. 2 e n. 6. La categoria A/Termi/DNT presenta valori che vanno da 12 a una sola occorrenza.

Esempio n. 5 (Prog1_Testo8 – A/Termi/DNT)

TO: Hintergaumenlaut [ɣ]

TA: suono del palato posteriore [ɣ]

PE: suono velare [ɣ]

Nell'esempio n. 5, la macchina tende a tradurre letteralmente il termine di partenza, senza riconoscerne l'equivalente corretto nel dominio di riferimento. La motivazione di tale tendenza può essere ricondotta all'assenza di testi specialistici nel dataset impiegato per il training del sistema.

Esempio n. 6 (Prog1_Testo1 – A/Termi/DNT)

TO: die Scanpaths der Probanden

TA: i percorsi di scansione dei soggetti in esame

PE: gli scanpath dei soggetti in esame

Anche in questo esempio si osserva come il sistema non sia in grado di individuare l'uso terminologico corretto del termine *scanpath*, che nel contesto specialistico non richiede traduzione.

Il sistema di TA tende invece a interpretarlo e tradurlo letteralmente, generando una resa impropria dal punto di vista terminologico.

In riferimento ai fenomeni linguistici del linguaggio scientifico, si registrano inoltre le seguenti occorrenze: 106 errori di F/SLP/Nomin, riguardanti l'impiego di nominalizzazioni (cfr. esempio n. 7) e 28 errori di F/SLP/ForImp, connessi all'impiego delle forme impersonali in sostituzione di quelle personali (cfr. esempio n. 8).

Esempio n. 7 (Prog1_Testo1 – F/SLP/Nomin)

TO: Nach der Kalibrierung des Eye-Trackers [...]

TA: Dopo aver calibrato l'eye tracker [...]

PE: Dopo la calibrazione dell'eye tracker [...]

Nell'esempio n. 7 la macchina non coglie la tendenza del linguaggio scientifico all'impiego di uno stile nominale, che conferisce al testo maggiore oggettività e compattezza stilistica. Il sistema di TA opta, invece, per una costruzione verbale esplicita, tipica del linguaggio comune.

Analogamente a quanto osservato nell'esempio n. 7, anche nell'esempio n. 8 la macchina non riconosce la convenzione stilistica del linguaggio scientifico, che privilegia forme impersonali per mantenere un tono neutro e distaccato.

Esempio n. 8 (Prog1_Testo6 – F/SLP/ForImp)

TO: Zur Beantwortung wird untersucht, ob [...]

TA: Per rispondere a questa domanda, analizzeremo se [...]

PE: Per rispondere a questa domanda, si analizzerà se [...]

Per quanto riguarda il lessico del linguaggio scientifico, sono stati riscontrati 45 errori di F/Lex/Fue, riferibili alla traduzione errata delle *Fügungen*, cioè le combinazioni idiomatiche composte da diversi lessemi tipiche del linguaggio scientifico (cfr. 4.3.3.2). L'esempio n. 9 illustra un errore di questa categoria, in cui la macchina traduce letteralmente la *Fügung* tedesca senza adattarla al linguaggio scientifico italiano.

Esempio n. 9 (Prog1_Testo4 – F/Lex/Fue)

TO: Das Resultat der Untersuchung von Yamamoto/Nakagawa (2006) lautet, dass [...]

TA: Il risultato dello studio di Yamamoto/Nakagawa (2006) è che [...]

PE: Dallo studio di Yamamoto/Nakagawa (2006) emerge che [...]

Infine, si segnalano 86 errori di F/Sty/Reg, che indicano l'impiego di un registro linguistico non adeguato al genere testuale (cfr. esempio n.10).

Esempio n. 10 (Prog1_Testo4 – F/Sty/Reg)

TO: Und laut Ogan/Aleven/Jones (2009) [...]

TA: E secondo Ogan/Aleven/Jones (2009), [...]

PE: Inoltre, secondo Ogan/Aleven/Jones (2009), [...]

Per quanto concerne la macrocategoria *Grammar*, questa include errori che incidono sulla correttezza della lingua di arrivo. Si osservano 173 errori relativi all'impiego della lettera maiuscola (G/Orth/Capit), 111 di punteggiatura (G/Orth/Punct), 81 di spaziatura (G/Orth/WhiSpa), 45 riguardanti l'accordo tra parole (G/WF/Agree), 39 di ortografia (G/Orth/Spel) e 6 relativi all'uso grammaticalmente scorretto di tempi o forme verbali (G/WF/VF). Pur non riguardando direttamente gli elementi caratterizzanti del linguaggio scientifico, tali errori risultano rilevanti per il raggiungimento di una qualità pubblicabile del testo. Di seguito sono riportati alcuni esempi dei casi che hanno registrato più occorrenze. L'esempio n. 11, tratto dal testo n. 9, illustra un errore tipico della categoria G/Orth/Capit (cfr. Tabella 17 – allegato 2).

Esempio n. 11 (Prog1_Testo9 – G/Orth/Capit)

TO: Die didaktische Umsetzung in diesem Lehrwerk lässt sich wie folgt zusammenfassen: Zunächst [...]

TA: L'implementazione didattica di questo libro di testo può essere riassunta come segue: In primo luogo,

PE: L'implementazione didattica di questo libro di testo può essere riassunta come segue: in primo luogo, [...]

In questo caso la traduzione automatica riflette l'uso ortografico tedesco della maiuscola dopo i due punti, quando segue una frase completa, senza adattarsi alla convenzione italiana che, diversamente, prevede la lettera minuscola.

L'esempio n.12 mostra un errore di G/Orth/Punct ed è tratto dal testo n. 4 che contiene un numero consistente di occorrenze (cfr. Tabella 1 – allegato 2).

Esempio n. 12 (Prog1_Testo4 – G/Orth/Punct)

TO: [...] intralinguale Untertitel –also Untertitel in der Fremdsprache – [...] begünstigen

TA: [...] che i sottotitoli intralinguistici –cioè i sottotitoli in lingua straniera –favoriscono [...]

PE: [...] che i sottotitoli intralinguistici₁ cioè i sottotitoli in lingua straniera₂ favoriscono [...]

In tedesco, gli incisi sono spesso segnalati con il *Gedankenstrich* (lineetta, –), segno molto frequente nella lingua scritta e in particolare nei testi scientifici. Il *Gedankenstrich* indica

principalmente una pausa più lunga tra parole o all'interno della frase. Può inoltre essere utilizzato per delimitare un segmento o un inciso e/o mettere in rilievo un'informazione, distinguendola chiaramente dal contesto sintattico in cui è inserita. Infine, può assumere un valore testuale o discorsivo, segnalando una transizione concettuale o tematica tra frasi o porzioni di testo (*Duden, Richtiges und gutes Deutsch*, 2001: 338). In questo caso, la lineetta viene impiegata in funzione parentetica, per introdurre un inciso esplicativo ("also Untertitel in der Fremdsprache"). Tale uso è tipico nei testi saggistici e scientifici, in cui il *Gedankenstrich* ha una funzione di disambiguazione e guida del processo di lettura (*leseprozessessteuerndes Zeichen*; cfr. Costa 2011: 176-183), segnalando un commento o una specificazione senza interrompere il flusso sintattico principale.

Nel sistema di interpunzione italiana, invece, la lineetta (o trattino lungo) ha un impiego molto più limitato (Lepschy e Lepschy 2008: 3-24). Gli incisi in lingua italiana sono infatti generalmente introdotti con la lineetta solo in alcuni casi, mentre più comunemente si ricorre a virgole, due punti, punto e virgola o parentesi, a seconda dei rapporti sintattico-semantiche all'interno della frase (cfr. Mortara Garavelli 2003: 105-108). Come osserva Costa (2011: 180), in italiano la coppia di trattini lunghi separa solitamente frasi subordinate o complementi con valore metacomunicativo, mentre nei casi in cui l'inciso abbia una funzione esplicativa o espletiva si preferisce l'uso delle virgole.

Alla luce di tali considerazioni, l'uso del trattino breve (-) della TA nell'esempio 12 è da considerarsi grammaticalmente errato: questo segno, infatti, viene utilizzato senza interruzioni di spazio per unire unità semantiche, segnalandone i rapporti sintattici (Mortara Garavelli 2003: 36-37). In fase di *post-editing*, la correzione è avvenuta mediante la sostituzione delle lineette con virgole, segno più conforme a delimitare un inciso esplicativo.

Significativa è anche la frequenza di alcune categorie di errori di *Accuracy*, che incidono sulla comprensibilità generale del testo e sulla fedeltà al testo di partenza: errori relativi a segmenti non tradotti (A/Untra, 213 occorrenze) o segmenti la cui traduzione non era necessaria (A/DNT, 198 occorrenze) ed errori di omissione di segmenti (A/Omis, 192 occorrenze). Gli errori di A/Untra segnalano i casi in cui il sistema di TA non è stato in grado di proporre una traduzione nella lingua di arrivo desiderata, riproponendo la frase nella lingua di partenza, come illustrato dall'esempio n. 13.

Esempio n. 13 (Prog1_Testo2 – A/Untra)

TO: Sie untersuchten die Äußerungen von 20 Vorschulkindern zwischen 3;11 und 5;3 Jahren.⁵⁸

TA: Sie untersuchten die Äußerungen von 20 Vorschulkindern zwischen 3;11 und 5;3 Jahren.

⁵⁸ Nel testo, l'età dei bambini è indicata nel formato "anni; mesi".

PE: Sono stati esaminati gli enunciati di 20 bambini in età prescolare compresa tra 3 anni e 11 mesi / 5 anni e 3 mesi.

Tale fenomeno rientra nel più generale problema dell'*under-translation*, tipico dei sistemi di TA neurale, in cui alcune parole o interi segmenti della lingua di partenza vengono erroneamente omessi o non tradotti (Shao et al. 2024). L'*under-translation* può verificarsi quando il modello incontra parole con alta *translation entropy* (entropia traduttiva), cioè termini che ammettono molte possibili traduzioni, o quando la funzione di *decoding* spinge il modello a privilegiare sequenze ad alta probabilità complessiva, trascurando segmenti ad alta incertezza traduttiva (Shao et al. 2024). Nel caso in esame, l'intervento del *post-editor* è stato necessario per inserire manualmente la traduzione corretta, restituendo completezza e coerenza al testo. Si registra un numero consistente di errori per questa categoria nel testo n.2, dove si riscontra questa problematica in 46 casi (cfr. Tabella 1 – allegato 2).

L'elevata frequenza della categoria A/DNT è attribuibile alla necessità di mantenere in lingua originale determinati enunciati o termini presi in esame nello studio, come menzionato nell'allegato 2 in riferimento al testo n. 2 e al testo n. 9. L'esempio n. 14, estratto dal testo n. 2, rientra in questa categoria, in quanto contiene un esempio linguistico impiegato nello studio originale in riferimento a un esercizio in tedesco somministrato a bambini che non deve essere tradotto.

Esempio n. 14 (Prog1_Testo2 – A/DNT)

TO: Nur ein Kind äußerte auf die Frage „Warum weint die Katze“ einen Hauptsatz [...]

TA: Solo un bambino ha risposto alla domanda "Perché il gatto piange" con una proposizione principale [...]

PE: Solo un bambino ha risposto alla domanda "Warum weint die Katze" con una proposizione principale [...]

La traduzione automatica tende a tradurre ogni segmento testuale senza distinguere tra contenuto comunicativo e contenuto metalinguistico. Al fine di preservare la coerenza metodologica e la validità scientifica del testo, in tali casi il *post-editor* deve identificare la frase come non appartenente al discorso dell'autore, ma come un esempio linguistico da mantenere nella lingua di partenza.

Un'ulteriore categoria di errori rimarchevole è A/ForLa che registra 162 occorrenze e che riguarda l'impiego di una lingua diversa da quella di arrivo e di partenza. Questa categoria è particolarmente rilevante per il testo n. 9, dove si registrano 80 casi di uso improprio della lingua

inglese, come illustra l'esempio n. 15 in riferimento all'uso dell'acronimo GFL per la traduzione in italiano dell'acronimo DaF.

Esempio n. 15 (Prog1_Testo9 – A/ForLa)

TO: [...] wie Jugendsprache in DaF-Lehrwerken behandelt wird [...]

TA: [...] come il linguaggio giovanile venga trattato nei libri di testo GFL [...]

PE: [...] come il linguaggio giovanile venga trattato nei libri di testo per il tedesco come lingua straniera [...]

In questo esempio, la macchina inserisce l'acronimo inglese "GFL" (*German as a Foreign Language*) al posto della traduzione italiana corretta di "DaF" (tedesco come lingua straniera), generando una discrepanza terminologica e linguistica. Questo fenomeno può essere spiegato considerando il funzionamento dei modelli di TA neurale: essi operano su sequenze di *token* e si basano sulle probabilità apprese dai dati di addestramento, nei quali l'inglese è sovrarappresentato. Pertanto, la lingua inglese funge da lingua *pivot*, cioè una lingua intermedia verso cui i *tokens* tendono a essere convertiti prima di essere espressi nella lingua di arrivo. Di conseguenza, il modello tende a preferire forme inglesi quando non riconosce un equivalente consolidato nella lingua di arrivo.

Altre categorie di *Accuracy* che condizionano il contenuto del testo di arrivo sono A/MisTra/WSD per gli errori di scelta semantica in riferimento a parole polisemiche (138 occorrenze) e A/MisTra/Part per gli errori riguardanti il trasferimento parziale o impreciso del significato (117 occorrenze). L'esempio n. 16 presenta un caso di errore di A/MisTra/WSD, mentre l'esempio n. 17 un caso per la categoria A/MisTra/Part. In entrambi i casi si tratta di interventi estremamente mirati.

Esempio n. 16 (Prog1_Testo1 – A/MisTra/WSD)

TO: [...] werden Reize ausgeblendet, die der Wahrnehmende als irrelevant erachtet [...]

TA: [...] gli stimoli che chi percepisce considera irrilevanti vengono sfumati [...]

PE: [...] gli stimoli che chi percepisce considera irrilevanti vengono disattivati [...]

Esempio n. 17 (Prog1_Testo4 – A/MisTra/Part)

TO: [...] mit Inhalten, die die Lernenden uninteressant finden [...]

TA: [...] con un contenuto che gli studenti trovano poco interessante [...]

PE: [...] con contenuti che gli studenti trovano poco interessanti [...]

Nell'esempio 17 la frase tradotta da TA è grammaticalmente corretta ma le due forme del numero grammaticale ("contenuto" / "interessante") non corrispondono a quelle del testo di partenza. Il *post-editing* ha ripristinato la resa corretta ("contenuti" / "interessanti").

Segue, in ordine decrescente di occorrenze, la categoria A/SourEr, che registra 84 occorrenze e che segnala gli errori presenti nel testo di partenza. Questi in alcuni casi possono riguardare la collocazione errata del numero della nota a piè di pagina all'interno della frase (cfr. esempio n. 18). Si tratta di un problema tecnico legato al file traducibile generato da Trados Studio: il numero inserito nel testo come rimando alla nota a piè di pagina viene posizionato in basso invece che in apice. Trattandosi di un errore presente già nel file di partenza generato da Trados, questo viene trasferito anche nel testo di arrivo, anche se non è presente nel file originale del testo di partenza.

Esempio n.18 (Prog1_Testo2 – A/SourEr)

TO (file originale): [...] der geäußerten Verben erfasst.³

TO (Trados studio): [...] der geäußerten Verben erfasst.³

TA: [...] di verbi pronunciati.³

PE: [...] di verbi pronunciati.³

In altri casi, gli errori di A/SouEr sono riconducibili alla formattazione non rispettata nel testo di partenza generato da Trados Studio e quindi anche nel testo di arrivo tradotto automaticamente. Tali fenomeni sono particolarmente evidenti nel testo n. 2, come evidenziato nella Tabella 3 dell'allegato 2. L'esempio n. 19 mostra un caso in cui l'uso del corsivo non è stato trasferito correttamente.

Esempio n.19 (Prog1_Testo4 – A/SouEr)

TO (file originale): (*critical thinking*)

TO (Trados studio): (*critical thinking*)

TA: (*critical thinking*)

PE: (*critical thinking*)

Si individuano ulteriori errori per alcune categorie di *Accuracy*, la cui presenza nei singoli testi è limitata. Tra questi figurano casi riguardanti la variazione di significato dovuta all'ordine delle parole (A/MisOrd/WO con 64 occorrenze). Tali errori non incidono direttamente sulle caratteristiche formali del linguaggio scientifico, ma compromettono la coerenza e la coesione del testo, ostacolandone la comprensione e la corretta trasmissione del contenuto. A titolo

esemplificativo viene riportato un caso per la categoria A/MisTra/WO, in cui la posizione differente della parola “anche” modifica il messaggio trasmesso dalla frase.

Esempio n.20 (Prog1_Testo2 – A/MisTra/WO)

TO: Auch Steel, Rose und Eadie (2013) stellten sich der Frage, ob Daten, [...]

TA: Steel, Rose e Eadie (2013) hanno anche affrontato la questione se i dati [...]

PE: Anche Steel, Rose e Eadie (2013) hanno affrontato la questione se i dati [...]

Si riscontra un numero rilevante di errori (64 occorrenze) per la categoria A/Add che segnala i segmenti aggiunti o ripetuti. L’esempio n. 21 illustra un caso per questa categoria, in cui l’abbreviazione tedesca “z. B.” è stata correttamente resa con “ad esempio”; tuttavia, la macchina ha inserito erroneamente l’elemento “B.”, trasferendolo impropriamente dal testo di partenza.

Esempio n. 21 (Prog1_Testo4 – A/Add)

TO: (z. B. Chagas 2005; Zhang 2013; Rouxel-Cubberly 2014).

TA: (ad esempio B. Chagas 2005; Zhang 2013; Rouxel-Cubberly 2014).

PE: (ad esempio Chagas 2005; Zhang 2013; Rouxel-Cubberly 2014).

Questo errore si verifica perché la traduzione automatica tratta ogni segmento come un’unità isolata, senza valutare la funzione complessiva dell’abbreviazione nel contesto. In altre parole, la macchina interpreta “B.” come un *token* autonomo da preservare, verosimilmente a causa della spazio all’interno dell’abbreviazione tedesca “z. B.”, che può indurre il sistema a trattare i due elementi come unità separate anziché come un’espressione abbreviativa unitaria; di conseguenza, non riconosce che nella lingua di arrivo la resa “ad esempio” ne esaurisce già il significato. Sebbene numericamente marginale, questo fenomeno assume particolare rilevanza in ambito scientifico, poiché può compromettere la correttezza delle citazioni bibliografiche, elemento essenziale per la validità e l’affidabilità del genere testuale considerato.

Proseguendo per numero decrescente di occorrenze, sono stati riscontrati 54 errori di A/MisTra/WroPr. Questa categoria evidenzia gli errori di preposizioni, sintagmi preposizionali o congiunzioni, come illustrato nell’esempio 22:

Esempio n. 22 (Prog1_Testo5 – A/MisTra/WroPr)

TO: [...] die Ergänzung und Weiterentwicklung durch die Dozentinnen und Dozenten vor Ort, die die Module einsetzen.

TA: [...] l’aggiunta e l’ulteriore sviluppo dei docenti in loco che utilizzano i moduli.

PE: [...] l’aggiunta e l’ulteriore sviluppo da parte dei docenti in loco che utilizzano i moduli.

Nell'esempio n. 22 la macchina ha proposto il sintagma preposizionale errato, rendendo la relazione tra il sintagma nominale "l'ulteriore sviluppo" e il complemento "i docenti" semanticamente ambigua.

Sono stati annotati inoltre 48 errori per la categoria A/MisExpl, riferibili all'inadeguatezza dei contenuti per il destinatario del testo di arrivo. L'esempio n. 23 ne mostra un caso.

Esempio n. 23 (Prog1_Testo6 – A/MisExpl)

TO: Der Satz »Herzlich willkommen!« soll [...] ausgesprochen werden.

TA: La frase "Benvenuto!" deve essere pronunciata [...]

PE: La frase "Herzlich willkommen!" (Benvenuto!) deve essere pronunciata [...]

La frase "Herzlich willkommen!" si riferisce a un esercizio presente in uno dei libri di testo analizzati nel testo n. 6. In fase di *post-editing*, l'inserimento dell'espressione tedesca tra virgolette e della traduzione italiana tra parentesi permette di mantenere la fedeltà al testo di partenza e di adattare il contenuto al destinatario italiano, garantendone la comprensione. Si osserva inoltre nel segmento un errore di F/Coh/InPu, relativo all'uso incoerente delle virgolette.

Tra le categorie di *Accuracy* che influiscono sul contenuto del testo e la sua corretta trasmissione, si individua A/MisTra/POS, che registra 45 occorrenze e segnala in particolare l'impiego di una categoria lessicale (*Wortart*) non corrispondente a quella del testo di partenza (cfr. esempio n. 24).

Esempio n. 24 (Prog1_Testo9 – A/MisTra/POS)

TO: [...] lässt sich ein hoher Anteil von medialer Mündlichkeit feststellen

TA: [...] si può osservare un'alta percentuale di media orali

PE: [...] si può osservare un'alta percentuale di oralità mediale

In questo esempio, il sistema di TA propone "media orali", invertendo la categoria lessicale dei due lessemi: l'aggettivo tedesco "medialer" viene tradotto come sostantivo "media", e il sostantivo "Mündlichkeit" come aggettivo "orali". Si tratta di un errore che compromette la correttezza concettuale del testo scientifico. Nel *post-editing*, la corretta traduzione, "oralità mediale", ristabilisce la precisione terminologica e il contenuto semantico del testo.

Un errore di particolare rilevanza per il linguaggio scientifico e per il genere testuale analizzato è rappresentato dalla categoria F/Coh/Cit che registra 41 occorrenze e che riguarda la strutturazione incoerente dei riferimenti bibliografici. L'esempio n. 25 riporta un caso relativo a questa categoria di errore, in cui la proposta del sistema di TA, pur corretta, risulta incoerente rispetto alle convenzioni adottate nel resto del testo. In particolare, poiché nei casi precedenti era

stato utilizzato “ibidem” come forma standard per le citazioni, la proposta automatica “ibid.” non rispetta l’uniformità di citazione all’interno del contributo.

Esempio n.25 (Prog1_Testo5 – F/Coh/Cit)

TO: (vgl. ebd.: 492).

TA: (cfr. ibid.: 492).

PE: (cfr. ibidem: 492)

Un’ulteriore categoria che conta un numero limitato di occorrenze, 32, è A/Mech/Punct che segnala gli errori relativi al trasferimento della punteggiatura. A titolo esemplificativo viene riportato un caso riguardante l’aggiunta di punteggiatura non necessaria nel segmento del testo di arrivo.

Esempio n. 26 (Prog1_Testo3 – A/Mech/Punct)

TO: (ebd.: 434).

TA: (ibid:).434).

PE: (ibid: 434).

Una categoria rilevante per la scientificità dell’articolo è A/MisTra/Gend, che segnala gli errori riguardanti la mancata resa del linguaggio inclusivo (28 occorrenze). Nell’esempio seguente, la traduzione automatica rende il termine generico “colleggi”, trascurando la distinzione di genere presente nel testo originale tedesco (“Kolleginnen und Kollegen”).

Esempio n. 27 (Prog1_Testo7 – A/MisTra/Gend)

TO: [...] Portfolio-Antworten von Kolleginnen und Kollegen [...]

TA: [...] risposte del Portfolio dei colleggi [...]

PE: [...] risposte del Portfolio delle colleghe e dei colleggi [...]

La correzione apportata in fase di *post-editing* riguardante l’esplicitazione di “colleghe e colleggi” consente il ripristino delle convenzioni linguistiche inclusive.

Tra le categorie di *Accuracy* che hanno un impatto sul livello semantico, si individuano i casi riferibili alla scorretta attribuzione di elementi (A/MisOrd/MW, 25 occorrenze). Nell’esempio n. 28 la proposta di traduzione automatica attribuisce in modo scorretto l’elemento “di ascolto”, collegandolo sia a “stili” sia a “strategie” e creando ambiguità, mentre il *post-editing* ristabilisce la corretta attribuzione, garantendo la trasmissione del contenuto corretto.

Esempio n. 28 (testo n. 8) (Prog1_Testo8 – A/MisOrd/MW)

TO: Übungstypen, Hörstile sowie Strategien.

TA: tipi di esercizi, stili e strategie di ascolto.

PE: tipi di esercizi, stili di ascolto e strategie.

Gli errori di F/Coh/CoRef riguardano la mancata corrispondenza tra entità citate in frasi diverse (21 occorrenze). Nonostante non incidano direttamente sugli aspetti formali del linguaggio scientifico, tali errori compromettono la coerenza e la coesione del testo, con ricadute negative sulla comprensione e la corretta trasmissione del contenuto. L'esempio n. 29 ne riporta un caso.

Esempio n. 29 (Prog1_Testo5 – F/Coh/CoRef)

TO: Auch diese wiederum kann allen Teilnehmenden [...] zur Verfügung gestellt werden.

TA: Questo può anche essere messo a disposizione di tutti i partecipanti [...].

PE: Questa può anche essere messa a disposizione di tutti i partecipanti [...].

Nell'esempio n. 29, il pronome dimostrativo “diese” si riferisce al sostantivo femminile “kommentierte Bibliographie” presente nella frase precedente. La traduzione automatica (“Questo può anche essere messo...”) neutralizza tale riferimento, sostituendo il femminile con un generico maschile (“questo”), formalmente concordato all'interno della frase (“messo”) ma non più coerente con l'antecedente, compromettendo così la coerenza testuale. La correzione effettuata in fase di *post-editing* ristabilisce il riferimento corretto, garantendo la comprensione e la coerenza.

Una categoria di errori che registra poche occorrenze (16 occorrenze) ma che risulta significativa per la correttezza formale del genere testuale del contributo scientifico è F/De/CharFor. Essa segnala gli errori relativi al mancato rispetto delle convenzioni tipografiche del testo di arrivo, che richiedono una formattazione diversa. Nell'esempio n. 30, il sistema di TA riporta correttamente le congiunzioni tedesche *weil*, *dass* e *wenn*, ma non applica la convenzione tipografica prevista per le parole straniere nel testo italiano, ossia l'uso del corsivo.

Esempio n. 30 (Prog1_Testo2 – F/De/CharFor)

TO: die durch die Konjunktionen weil, dass und wenn eingeleitet werden können.

TA: che possono essere introdotte dalle congiunzioni weil, dass e wenn.

PE: che possono essere introdotte dalle congiunzioni weil, dass e wenn.

Infine, le categorie meno rappresentate sono le seguenti: F/De/Layo con 11 occorrenze; A/Mech/CharFor con 9 occorrenze; A/MWE/Ell e A/MWE/Komp con 8 occorrenze; F/Sty/Idiom con 6 occorrenze; A/Mech/Capit con 4 occorrenze. Tali dati confermano che gli errori relativi alla strutturazione dei paragrafi e del testo di arrivo (F/De/Layo), al trasferimento della formattazione

(A/Mech/CharFor) e delle lettere maiuscole (A/Mech/Capit), alla traduzione di aggettivi composti (A/MWE/Komp) o di parole composte con ellissi di elementi ripetuti (A/MWE/Ell) e alla resa di espressioni idiomatiche (F/Sty/Idiom) risultano marginali in questo progetto.

Nel complesso, gli errori direttamente riconducibili, per definizione delle categorie di appartenenza, al linguaggio scientifico ammontano a 1.403, poco più di un quinto del totale (circa 21,5%). Tali categorie, relative ad esempio a nominalizzazioni, forme impersonali, terminologia specialistica o strutturazione delle citazioni, sono state definite appositamente per gli elementi tipici del linguaggio scientifico e richiedono un intervento umano mirato per garantire una resa in lingua italiana che sia coerente e pubblicabile nel contesto di questo progetto di traduzione. Le altre tipologie di errore, pur non afferendo direttamente al linguaggio scientifico, influenzano in modo significativo la qualità complessiva del testo e la resa del linguaggio scientifico stesso e dei suoi elementi: gli errori di *Accuracy* incidono sulla corretta trasmissione del contenuto, quelli di *Fluency* sulla scorrevolezza della lingua di arrivo e quelli di *Grammar* sulla correttezza formale. Pertanto, anche gli interventi di correzione relativi a queste categorie hanno contribuito in maniera determinante al raggiungimento di un prodotto finale che rispetti le convenzioni del linguaggio scientifico e garantisca accuratezza e coerenza.

6.2.2 Analisi dell'effort temporale e tecnico

Il totale delle sessioni per il progetto di traduzione *Informationen Deutsch als Fremdsprache* ammonta a 120:29:08⁵⁹, con una durata media per testo di 13:24:09. Le sessioni di una durata superiore a 01:00:00 circa sono state intervallate con pause ogni 01:30:00/01:50:00, escluse dal conteggio della durata delle sessioni⁶⁰.

Il report di Post-Edit Compare relativo a questo progetto di traduzione prende in esame complessivamente 2.762 segmenti e circa 52.232 parole. La quasi totalità del materiale (2.679 segmenti) è stata tradotta dal sistema di TA DeepL, mentre solo 83 segmenti figurano come traduzione non proposta dal sistema di TA e inserita manualmente.

È opportuno precisare che Post-Edit Compare opera a livello di progetto di traduzione e richiede una versione di partenza unica per tutti i testi. Per questa ragione, la versione iniziale è stata salvata prima dell'avvio del *post-editing*, quando la memoria di traduzione, creata per il progetto, era ancora vuota. Sebbene la memoria si sia arricchita progressivamente durante la correzione, tale aggiornamento non influisce sull'analisi, poiché lo strumento lavora solo sulla versione iniziale salvata, senza possibilità di ricalcolare eventuali *matches* introdotti

⁵⁹ Tutte le durate riportate in questo capitolo sono espresse nel formato ore:minuti:secondi (hh:mm:ss).

⁶⁰ Le indicazioni temporali per ogni singolo testo sono riportate nell'allegato 2.

successivamente. L'assenza di contributi provenienti dalla memoria di traduzione nel report, dunque, non indica un'anomalia, bensì riflette il funzionamento dello strumento e conferma che l'intero testo di partenza valutato da Post-Edit Compare è costituito unicamente dall'output della TA.

La Tabella 5 illustra le percentuali complessive rilevate nelle diverse fasce di PEMp, riportando per ciascuna di esse il numero associato di segmenti e di parole, al fine di offrire una visione dettagliata della distribuzione dei dati.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	591	7.669	14,68%
95-99%	302	6.470	12,38%
85-94%	614	15.095	28,92%
75-84%	348	7.981	15,29%
50-74%	276	5.124	9,82%
New (0-49)	631	9.893	18,91%

Tabella 5: Distribuzione delle fasce di PEMp per progetto n. 1

La distribuzione delle modifiche evidenzia un andamento eterogeneo, pur mostrando alcune tendenze significative. Una parte non trascurabile dei segmenti è stata accettata senza variazioni: la fascia 100% raccoglie il 14,68% delle parole (591 segmenti, 7.669 parole), indicando che circa un settimo del testo fornito dalla TA era già pronta all'uso, senza intervento da parte del *post-editor*. Nella Figura 29 viene riportato un segmento estratto dal report di Post-Edit Compare, che costituisce un caso per questa fascia di PEMp. Si può osservare il segmento di partenza per il quale il sistema di TA ha proposto una traduzione che è stata accettata senza modifiche, per cui la casella del *post-editing* risulta vuota nel report. Nell'ultima colonna a destra viene riportata la percentuale di PEMp, in questo caso 100%, con un *edit-distance* pari a 0, a conferma che non sono state apportate modifiche.

Testo di partenza	Traduzione automatica	Post-editing	PEMp
Im Folgenden werden einige Beispiele genannt:	Di seguito sono riportati alcuni esempi:		100% Edit-Dist.: 0 Max chars: 40

Figura 29: Esempio fascia di PEMp 100% dal testo n. 8 del progetto n. 1

Secondo la Tabella 5, la fascia 95-99% presenta il 12,38% (302 segmenti, 6.470 parole), segnalando i segmenti che hanno richiesto solo correzioni minime. Tale dato indica una diffusione contenuta di questa tipologia di interventi, la cui quota più consistente si concentra nella fascia 85-94%, che da sola costituisce il 28,92% delle parole (614 segmenti, 15.095 parole). Ciò risulta in linea con quanto emerge nei report riguardanti i singoli testi presentati nell'allegato 2 in cui questa

fascia di PEMp è sempre la più consistente con un'unica eccezione rappresentata dal testo n. 6. Tale percentuale è particolarmente significativa, in quanto indica che quasi un terzo del testo ha richiesto interventi di *post-editing* di lieve entità. Ciò significa per il *post-editor* agire mediante interventi mirati, come mostra la Figura 30. Nella colonna “Confronto” si possono osservare le modifiche effettuate, evidenziate in giallo e scritte in blu, rispetto al testo proposto dalla macchina in seguito alla traduzione automatica, barrato in rosso. La percentuale di PEM del segmento è 90,91%.

Testo di partenza			
Bei der Analyse der drei Lehrwerke wurde deutlich, dass in den Lehrwerken viele Strategien zumindest implizit enthalten sind.			
Traduzione automatica	Post-editing	Confronto	PEMp
Analizzando i tre libri di testo, è emerso chiaramente che molte strategie sono incluse almeno implicitamente nei libri di testo.	Dall'analisi dei tre libri di testo, è emerso chiaramente che molte strategie sono incluse almeno implicitamente nei libri di testo.	Analizzando Dall'analisi dei tre libri di testo, è emerso chiaramente che molte strategie sono incluse almeno implicitamente nei libri di testo.	90.91% Edit-Dist.: 12 Max chars: 132

Figura 30: Esempio fascia di PEMp 85-94% dal testo n. 8 del progetto n. 1

Proseguendo con l'analisi delle fasce inferiori, si osservano la fascia 75-84% con il 15,29% (348 segmenti, 7.981 parole) e la fascia 50-74%, che raggiunge il 9,82% (276 segmenti, 5.124 parole), indicative di interventi di media e medio-alta entità, questi ultimi di densità particolarmente ridotta. Infine, la categoria *New* copre il 18,91% delle parole (631 segmenti, 9.893 parole), segnalando che una porzione consistente di segmenti ha richiesto riscritture integrali o pressoché integrali, cioè modifiche in una percentuale che va da 49% a 0%. La Figura 31 mostra un segmento interessato da una percentuale di parole rimaste invariate a seguito del *post-editing* del 23,98%. Il segmento proposto dalla TA ha subito una rielaborazione consistente, come si può osservare nella colonna dedicata al confronto tra la versione automatica e quella post-editata.

Testo di partenza
Für das Hörverstehen der deutschen Sprache lassen sich einige der Ergebnisse in analoger Übertragung anwenden.

Traduzione automatica	Post-editing	Confronto	PEMp
Per la comprensione orale del tedesco, alcuni dei risultati possono essere applicati al trasferimento analogico.	Alcuni di questi risultati possono essere applicati in modo analogo alla comprensione orale della lingua tedesca.	PerAlcuni ladi comprensionequesti oralerisultati delpossono tedesco;essere alcuniapplicati deiln risultatimodo pessoneanalogo esserealla applicaticomprendione alorale trasferimentodella analogicoilingua tedesca.	23.89% Edit-Dist.: 86 Max chars: 113

Figura 31: Esempio categoria di PEMp *New* dal testo n.8 del progetto n. 1

Complessivamente emerge che solo una parte limitata del testo presentava già una qualità pubblicabile senza necessitare ulteriori interventi, indice di una scarsa affidabilità del testo prodotto automaticamente. La restante porzione è stata invece oggetto di interventi di *post-editing* di entità variabile: circa un terzo di essa presenta revisioni di media entità, mentre i rimanenti due terzi comprendono interventi di diversa intensità, tra cui le riscritture totali o quasi totali, che rappresentano una quota considerevole. In termini di *effort*, è stato affrontato un lavoro di *post-editing* disomogeneo, alternando segmenti pronti all'uso ad altri che hanno richiesto modifiche approfondite o rielaborazioni complete. Il report mostra chiaramente che in questo contesto, sebbene la TA fornisca un supporto consistente su testi e progetti di traduzione di dimensioni elevate, il suo contributo non è uniforme. Può dunque essere considerata un punto di partenza, tenendo presente che la qualità varia notevolmente da segmento a segmento all'interno dei testi stessi e che il *post-editor* deve modulare il proprio intervento in base alla complessità delle correzioni necessarie.

6.3 Progetto n. 2: *trans-kom*

In questo paragrafo vengono presentate le analisi dei dati complessivi per i testi inclusi nel progetto di traduzione *trans-kom*. Il numero di *tokens* complessivo del progetto ammonta a 80.711.

6.3.1 Analisi degli errori⁶¹

Categoria di errore	Somma delle annotazioni
F/Coh/InPu	1713
A/MisTra/SemUn	750

⁶¹ Le analisi dei singoli testi sono state riportate nell'allegato 3.

F/Sty/Disfl	742
F/Lex/LexCho	537
A/DNT	335
A/Termi/SemUn	314
A/Untra	224
A/MisTra/WSD	205
A/Omis	191
G/Orth/Capit	189
F/Sty/Reg	151
G/Orth/Punct	120
A/SouEr	116
G/Orth/WhiSpa	114
A/Mech/Punct	114
F/De/CharFor	92
A/MisTra/Part	90
F/SLP/Nomin	75
A/ForLa	65
A/MisOrd/WO	61
F/SLP/ForImp	59
A/MisExpl	50
A/Termi/DNT	49
F/Lex/Fue	48
F/De/Layo	47
G/WF/Agree	44
A/MisTra/Gend	44
G/Orth/Spel	43
A/MisTra/WroPr	41
A/MisTra/POS	39
F/Coh/CoRef	38
A/Mech/CharFor	35
A/Add	32
F/Coh/Cit	29
A/Oth	20
A/MWE/Komp	17
F/Sty/Idiom	15
G/WF/VF	15
A/Mech/Capit	13
A/MisOrd/MW	11
A/MWE/Ell	6

Tabella 6: Analisi quantitativa degli errori per progetto n. 2

Nel complesso, il progetto di traduzione evidenzia un totale di 6.893 errori, suddivisi in 3.546 errori di *Fluency*, 2.822 errori di *Accuracy* e 525 errori di *Grammar*. La categoria quantitativamente più considerevole è la *Fluency*, principalmente a causa della rilevante incidenza degli errori di F/Coh/InPu, che contano 1.713 occorrenze. Gli errori appartenenti a questa categoria

segnalano un uso incoerente di segni ortografici, in particolare l'apostrofo e le virgolette (cfr. esempio n. 31).

Esempio n. 31 (Prog2_Testo5 – F/Coh/InPu)

TA: [...] sono "diversi" [...]

PE: [...] sono “diversi” [...]

Come evidenziano le analisi quantitative dei singoli testi riportate nell'allegato 3, questa categoria è la più frequente, con le sole eccezioni dei testi n. 4 e n. 7. Tuttavia, in accordo con quanto affermato nell'analisi complessiva precedente, tali errori non hanno ripercussioni dirette sulla qualità della traduzione del linguaggio scientifico: si tratta di fenomeni riscontrabili in maniera trasversale in qualsiasi dominio, genere testuale e coppia linguistica sottoposta alla TA.

Seguendo l'ordine decrescente di frequenza, si registrano 750 errori di A/MisTra/SemUn e 741 errori di F/Sty/Disfl. Gli errori di A/MisTra/SemUn indicano la mancata corrispondenza semantica tra testo di partenza e testo di arrivo; diversamente, gli errori di F/Sty/Disfl segnalano problematiche di scorrevolezza e naturalezza della lingua di arrivo. Alla luce di queste informazioni, i casi appartenenti a queste due categorie sono rilevanti per la comprensione del testo di arrivo e la fedeltà al testo di partenza. Inoltre, queste categorie rientrano tra quelle maggiormente annotate nei testi sottoposti a *post-editing*, collocandosi nelle tabelle quantitative degli errori illustrate nell'allegato 3 tra le posizioni uno e quattro. Di seguito sono riportati due esempi: l'esempio n. 32 per la categoria A/MisTra/SemUn e l'esempio n. 33 per la categoria F/Sty/Disfl.

Esempio n.32 (Prog2_Testo5 – A/MisTra/SemUn)

TO: [...], wie schon von Schreiber (1993: 76) verlangt, [...]

TA: [...], come già descritto da Schreiber (1993: 76), [...]

PE: [...], come già richiesto da Schreiber (1993: 76), [...]

Esempio n.33 (Prog2_Testo5 – F/Sty/Disfl)

TO: Solche Übersetzungen sollen die Werte der Zielkultur einem gewissen Druck aussetzen [...]

TA: Tali traduzioni hanno lo scopo di esporre i valori della cultura d'arrivo a una certa pressione [...]

PE: Tali traduzioni hanno lo scopo di sottoporre i valori della cultura d'arrivo a una certa pressione [...]

Un'ulteriore categoria che registra un numero consistente di errori è A/DNT con 335 occorrenze. Si tratta nello specifico di segmenti, una o più parole, la cui traduzione non era

necessaria. Il numero elevato di occorrenze per questa categoria è dovuto alle tematiche degli articoli: in alcuni casi si constata la presenza di diverse citazioni in lingue straniere, tra le quali il francese e l'inglese. In fase di *post-editing* si è deciso di rispettare la scelta dell'autore non traducendo tali citazioni, che erano state tradotte dal sistema di TA, com'è accaduto nel testo n. 5 e n. 6 (cfr. tabelle 9 e 11 – allegato 3). Di seguito si riporta l'esempio di un segmento tratto da una citazione infratesto:

Esempio n. 34 (Prog2_Testo3 – A/DNT)

TO: In this definition, 'compossibility' means capability for co-existence and interaction. (Dolezel 1995: 94)

TA: In questa definizione, 'compossibilità' significa capacità di coesistenza e interazione (Dolezel 1995: 94)

PE: In this definition, 'compossibility' means capability for co-existence and interaction. (Dolezel 1995: 94)

In altri casi, gli articoli presentano studi di traduzione su parole specifiche. Si tratta di lessemi utilizzati come oggetto di studio, citati per riflettere su aspetti linguistici e terminologici. Se tradotti, il significato e la funzionalità del testo di partenza vengono alterati. Questo è quanto avviene nel testo n. 8, di cui un errore viene illustrato nell'esempio n. 35.

Esempio n. 35 (Prog2_Testo8 – A/DNT)

TO: [...] beim Kompositum Moorbad [...], das mit 'bagno di fango, fanghi' [...] übersetzt wird,

TA: [...] nel composto bagno di fango, che viene tradotto come 'bagno di fango, fanghi' [...]

PE: [...] nel composto Moorbad, che viene tradotto come 'bagno di fango, fanghi' [...]

Un ulteriore caso di errore di A/DNT è tratto dal testo n. 9, che presenta un numero elevato di occorrenze per questa categoria a causa di numerosi esempi presi in esame nello studio volto a osservare l'uso degli articoli in lingua italiana e tedesca. Di seguito viene mostrato un esempio:

Esempio n. 36 (Prog2_Testo9 – A/DNT)

TO: Bsp. (3)

Wir haben Mehl.

TA: Esempio (3)

Abbiamo la farina.

PE: Esempio (3)

Wir haben Mehl.

Questi esempi mostrano come gli errori di A/DNT derivino dalla tendenza del sistema di TA a tradurre segmenti che, per motivi di contenuto o di funzionalità scientifica, devono rimanere nella lingua originale. L'intervento del *post-editor* risulta quindi essenziale per preservare la fedeltà, il significato e la funzione didattica o terminologica del testo, garantendo che citazioni, termini specialistici ed esempi di studio mantengano la loro integrità all'interno del documento.

Particolare rilievo assume invece la categoria F/Lex/LexCho, che registra 538 errori e che concerne l'uso di espressioni appropriate all'ambito specialistico e al genere testuale trattati, come mostra l'esempio n. 37. Questa categoria rientra tra quelle che registrano più risultati collocandosi in una posizione variabile tra il terzo e il quinto posto nelle tabelle quantitative degli errori illustrate nell'allegato 3.

Esempio n. 37 (Prog2_Testo6 – F/Lex/LexCho)

TO: Allerdings hat dieses Schreiben einen sowohl klar definierten Ausgangspunkt/-text als auch einen fest gelegten Endpunkt/-text.

TA: Tuttavia, questa scrittura ha sia un punto/testo iniziale che un punto/testo finale ben definiti.

PE: Tuttavia, questa scrittura ha sia un punto/testo di partenza che un punto/testo di arrivo ben definiti.

Per quanto riguarda la terminologia, si rilevano 314 errori di A/Termi/SemUn, riferibili alla traduzione non corretta di termini specialistici (cfr. esempio n. 38), e 49 errori di A/Termi/DNT, categoria destinata a segnalare i termini che non richiedono traduzione (cfr. esempio n. 39). Tali categorie risultano particolarmente rilevanti nei contributi scientifici, in quanto la terminologia costituisce un elemento distintivo del linguaggio scientifico e contribuisce in modo decisivo alla chiarezza e all'autorevolezza del discorso.

Esempio n. 38 (Prog2_Testo2 – A/Termi/SemUn)

TO: [...] Vorschläge der Skopos-Theorie [...].

TA: [...] le proposte della teoria di Skopos [...].

PE: [...] le proposte della teoria dello Skopos [...].

Nel caso dell'esempio n. 38, l'errore riguarda la traduzione errata del termine "Skopos-Theorie", reso come "teoria di Skopos" invece di "teoria dello Skopos". Quest'ultima è la formulazione comunemente accettata nella letteratura traduttologica in lingua italiana, rispettando la funzione del termine "Skopos" come concetto teorico e non come nome proprio.

Esempio n.39 (Prog2_Testo1 – A/Termi/DNT)

TO: [...] Travelling-Concepts-Ansatzes [...]

TA: [...] dell'approccio dei concetti itineranti [...]

PE: [...] dell'approccio dei travelling concepts [...]

Nell'esempio n. 39, il termine "travelling concepts" non viene tradotto poiché è stabilmente utilizzato nella letteratura specialistica in lingua italiana, dove mantiene la forma inglese per garantire precisione terminologica e coerenza con l'uso accademico internazionale consolidato.

Proseguendo nell'analisi degli aspetti caratteristici del linguaggio scientifico, si evidenziano 151 errori di F/Sty/Reg, legati all'inadeguatezza del registro linguistico. Questa categoria risulta particolarmente rilevante per il testo n. 1, che registra 55 occorrenze (cfr. allegato 3). L'esempio n. 40 riporta un caso in cui l'errore di F/Sty/Reg ha coinvolto più parole.

Esempio n.40 (Prog2_Testo6 – F/Sty/Reg)

TO: Wenn im übersetzten Text nicht mehr an-, sondern abgeschraubt steht, [...]

TA: Se il testo tradotto non dice più vitato ma svitato, [...]

PE: Se nel testo tradotto non compare più vitato ma svitato, [...]

In riferimento ai fenomeni linguistici del linguaggio scientifico, si registrano le seguenti occorrenze: 75 errori di F/SLP/Nomin, riguardanti l'impiego di nominalizzazioni (cfr. esempio n. 41) e 59 errori di F/SLP/ForImp, connessi all'impiego delle forme impersonali in sostituzione di quelle personali (cfr. esempio n. 42). Di seguito sono riportati due esempi:

Esempio n. 41 (Prog2_Testo5 – F/SLP/Nomin)

TO: eine Befindlichkeit, die bei der Konfrontation mit [...] entstehen kann.

TA: uno stato d'animo che può sorgere quando ci si confronta con [...]

PE: uno stato d'animo che può sorgere nel confronto con [...]

In questo esempio, la traduzione automatica non rispetta la tendenza del linguaggio scientifico a privilegiare strutture nominali, preferendo invece una costruzione verbale. L'impiego del sintagma nominale ("nel confronto con") nella versione post-editata contribuisce a rendere il testo più compatto e impersonale, in linea con lo stile informativo e oggettivo tipico del contributo scientifico.

Esempio n.42 (Prog2_Testo4 – F/SLP/ForImp)

TO: Betrachtet man die potenziellen Zielgruppen [...] nun jedoch genauer [...]

TA: Tuttavia, se osserviamo più da vicino i potenziali destinatari [...]

PE: Tuttavia, se si osservano più da vicino i potenziali destinatari [...]

Nell'esempio n. 42 l'errore riguarda la mancata resa della costruzione impersonale, caratteristica del linguaggio scientifico, che mira a mantenere un tono neutro e distaccato. La forma personale ("se osserviamo") adottata dalla traduzione automatica introduce invece un punto di vista soggettivo, inadeguato al registro scientifico.

Inoltre, sono stati riscontrati 48 errori di F/Lex/Fue, riferibili alla traduzione errata delle *Fügungen* e pertanto particolarmente significativi per la traduzione del linguaggio scientifico tedesco in italiano. L'esempio n. 43 ne illustra un caso.

Esempio n. 43 (Prog2_Testo4 – F/Lex/Fue)

TO: Der vorliegende Beitrag soll die Herausforderungen thematisieren, mit denen [...]

TA: L'obiettivo di questo articolo è quello di affrontare le sfide che [...]

PE: Il presente contributo intende mettere in evidenza le sfide con cui [...]

In questo caso, il sistema di TA rende "Der vorliegende Beitrag soll...thematisieren" con "L'obiettivo di questo articolo è quello di affrontare...", optando per una costruzione più vicina al linguaggio comune. Nella versione post-editata, invece, si adotta la formulazione "Il presente contributo intende mettere in evidenza...", che rende in modo più adeguato il valore semantico di "thematisieren" e rispetta le convenzioni del linguaggio scientifico italiano. Tale scelta mantiene un registro formale e preserva la densità informativa della *Fügung* originale. Questo tipo di errore è particolarmente rilevante in testi scientifici, poiché una traduzione letterale o inadeguata di combinazioni lessicali consolidate può influire sulla precisione concettuale e sulla leggibilità formale del testo.

La categoria degli errori grammaticali (*Grammar*) comprende casi che non influenzano direttamente il linguaggio scientifico, ma incidono sulla correttezza della lingua di arrivo. Tra questi si contano: 189 errori relativi all'impiego della lettera maiuscola (G/Orth/Capit), 120 errori di punteggiatura (G/Orth/Punct), 114 errori di spaziatura (G/Orth/WhiSpa), 44 errori di accordo tra parole (G/WF/Agree), 43 errori di ortografia (G/Orth/Spel) e 15 errori riguardanti l'uso di tempi e forme verbali (G/WF/VF). Per gli errori di G/Orth/Punct si constata la stessa problematica illustrata nell'esempio n. 12 dell'analisi complessiva precedente che riguarda la punteggiatura per gli incisi. Particolarmente considerevole è la categoria G/Orth/Capit nel testo n. 9 che registra 43 errori (cfr. Tabella 17 – allegato 3).

L'esempio n. 44 presenta un caso relativo alla categoria G/Orth/WhiSpa, il cui numero di occorrenze è rilevante nel testo n. 1 (cfr. Tabella 1 – allegato 3). Nell'esempio riportato si verifica l'assenza di spaziatura tra il punto fermo e la lettera maiuscola che segna l'inizio di una nuova frase.

Esempio n. 44 (Prog2_Testo1 – G/Orth/WhiSpa)

TA: ed è quindi soggetto a tutti i suoi limiti. Nonostante tutti questi [...]

PE: ed è quindi soggetto a tutti i suoi limiti. Nonostante tutti questi [...]

Significativa è anche la frequenza di alcune categorie di errori di *Accuracy*, che compromettono la comprensibilità generale e la fedeltà al testo di partenza. Queste riguardano segmenti non tradotti (A/Untra, 224 errori) o omessi (A/Omis, 191 errori) e la scelta semantica errata per parole polisemiche (A/MisTra/WSD, 205 errori). La categoria A/Untra si distingue per la sua rilevanza nel testo n. 2 che, come emerge dalla Tabella 3 dell'allegato 3, registra 61 occorrenze interessando singole parole, sintagmi nominali (cfr. esempio 45) e frasi intere.

Esempio n. 45 (Prog2_Testo2 – A/Untra)

TO: 6 Mehrdimensionale Beurteilung: Zusammenfassung

TA: 6 Mehrdimensionale Beurteilung: sintesi

PE: 6 Valutazione multidimensionale: sintesi

La mancata traduzione del segmento può essere ricondotta al fenomeno di *under-translation*, già menzionato in riferimento all'esempio n. 13 del progetto n. 1. Il sistema di TA ha mantenuto il segmento nella lingua di partenza, non riuscendo a generare la corrispondente traduzione nella lingua di arrivo.

Dalla tabella quantitativa degli errori riportata nell'allegato 3 per il testo n.1, emerge che la categoria di errori A/Omis è particolarmente significativa. Con un totale di 45 errori, essa evidenzia numerosi casi di omissione di segmenti, che possono riguardare singole parole, parti di frasi o frasi intere. Un esempio di tale fenomeno è riportato nell'esempio n. 46.

Esempio n. 46 (Prog2_Testo1 – A/Omis)

TO: “Mindestens müssen sich nämlich [um die Differenz von Kulturen überhaupt bemerken zu können, TK] die sprachliche Bezugnahme auf eine andere Kultur als ‘Objekt’ der Untersuchung und der praktische Zugang zu diesem Objekt unterscheiden.” (Renn 2005: 199).

TA: (Renn 2005: 199).

PE: “Infatti, [per poter notare la differenza tra culture, aggiunta] è necessario che si distinguano il riferimento linguistico a un'altra cultura in quanto ‘oggetto’ dell'indagine e l'accesso pratico a questo oggetto.” (Renn 2005: 199).

Nell'esempio n. 46, anche l'omissione quasi integrale della frase è riconducibile al fenomeno di *under-translation*. Secondo Yang et al. (2019), le omissioni rappresentano un problema sistemico

dei motori di TA neurale ampiamente diffuso, legato ai meccanismi di apprendimento e di *decoding*.

Come emerge dalla Tabella 3 riportata nell'allegato 3, la categoria A/MisTra/WSD assume un ruolo considerevole nel testo n.1 con 79 occorrenze. Di seguito viene riportato un esempio, in cui è stato realizzato un intervento di *post-editing* mirato.

Esempio n.47 (Prog2_Testo1 – A/MisTra/WSD)

TO: Sinn [...] kann ohne die von Reinart als subjektiv abqualifizierten Faktoren gar nicht entstehen.

TA: il senso [...] non può crearsi senza i fattori che Reinart squalifica come soggettivi.

PE: il senso [...] non può crearsi senza i fattori che Reinart dequalifica come soggettivi.

Nell'esempio n. 47, l'intervento di *post-editing* ha corretto la discrepanza semantica riguardante la traduzione del lessema "abqualifizierten", che comprometteva la fedeltà al testo di partenza e l'accuratezza concettuale.

Un'ulteriore categoria rilevante di *Accuracy* è A/SouEr che registra 116 occorrenze e riguarda gli errori presenti nel testo di partenza. Gli errori della categoria A/SouEr sono prevalentemente determinati dalla collocazione errata del numero della nota a piè di pagina nel file traducibile del testo di partenza creato da Trados Studio, come già segnalato nell'analisi complessiva precedente (cfr. esempio n. 18 in 6.2.1). L'analisi dei dati riportati nella Tabella 17 dell'allegato 3 mostra che si tratta di un errore particolarmente significativo per il testo n. 9.

La categoria A/Mech/Punct, che segnala il trasferimento errato della punteggiatura, registra 114 risultati (cfr. esempio n. 26 in 6.2.1). Si tratta di una categoria di errore particolarmente ricorrente nel testo n. 8, come mostrato dalla Tabella n. 15 nell'allegato 3.

Un numero consistente di occorrenze riguarda successivamente la categoria di errori F/De/CharFor, che include gli errori di formattazione e di convenzioni tipografiche del testo di arrivo (92 errori). Significativo è il numero di occorrenze per questa categoria nel testo n. 4, come emerge dalla tabella quantitativa degli errori presente nell'allegato 3. L'esempio n. 48 mostra un caso per questa categoria estratto da questo testo, in cui l'errore di F/De/CharFor segnala la necessità di scrivere in corsivo una parola in una lingua diversa da quella di arrivo.

Esempio n. 48 (Prog2_Testo4 – F/De/CharFor)

TO: In Leichte-Sprache-Texten [...]

TA: Nei testi in linguaggio semplice [...]

PE: Nei testi in *leichte Sprache* [...]

In questo caso, l'errore di F/De/CharFor è accompagnato da un errore di A/Termini/DNT, dovuto alla necessità di non tradurre il termine. Nel testo di partenza, *Leichte Sprache* designa una varietà linguistica del tedesco creata per rendere i testi scritti accessibili a persone con difficoltà di apprendimento, sulla base di regole e linee guida codificate (Crestani 2024: 200-201). Essa si distingue dalla *Einfache Sprache*, che corrisponde alla *plain language*, intesa da Perego (2021: 280) come la varietà destinata a un pubblico generalista che accede a contenuti specialistici. Crestani (2022: 242) propone come equivalente italiano di *plain language*, e quindi di *einfache Sprache*, la denominazione “linguaggio semplice”. Pertanto, la proposta di traduzione del sistema di TA risulta inadeguata.

Come osserva Crestani (2022: 243), la lingua italiana presenta una notevole variabilità terminologica per indicare la varietà della *leichte Sprache*: “lingua facile”, “linguaggio facile”, “linguaggio facile da leggere e da capire”, “linguaggio semplice da leggere e da capire”, o “italiano facile”, con oscillazioni di uso anche diatopiche (ad esempio, “lingua facile” e “linguaggio facile” in Alto Adige e Svizzera). In ambito accademico italiano, il termine “linguaggio facile” è impiegato per rendere *Leichte Sprache*, infatti Crestani (2022) propone il termine “linguaggio facile italiano”, pur sottolineando che “questo non sempre rispecchia la categorizzazione fornita nei testi prodotti in questa varietà” (Crestani 2024: 201). La stessa autrice utilizza le denominazioni *Deutsche Leichte Sprache* e “linguaggio facile italiano” per riferirsi rispettivamente alla varietà tedesca e a quella italiana (Crestani 2024).

Alla luce di queste considerazioni, in fase di *post-editing* si è deciso di mantenere il termine tedesco *Leichte Sprache*. Tale scelta risponde a diverse esigenze traduttive e comunicative. In primo luogo, in italiano non esiste un equivalente univoco e codificato che restituisca pienamente il significato e le implicazioni sociolinguistiche del termine tedesco. Inoltre, *Leichte Sprache* è un concetto fortemente radicato nel contesto culturale e istituzionale tedesco, legato a pratiche normative e linee guida specifiche, e una sua traduzione rischierebbe di attenuarne la portata. Infine, poiché il testo è rivolto a un pubblico specializzato, il prestito identico consente di garantire una maggiore precisione terminologica e di preservare la coerenza con la terminologia usata nella letteratura scientifica di riferimento. Conseguentemente alla scelta di usare il termine in tedesco anche nel testo di arrivo, esso è stato riportato in corsivo. L'intervento umano in fase di *post-editing* evidenzia quindi la capacità di riconoscere sfumature linguistiche e socioculturali che la traduzione automatica tende a neutralizzare, restituendo così un testo terminologicamente più accurato e coerente con il registro scientifico.

In ordine decrescente di occorrenza, si riscontrano inoltre le seguenti categorie di errori: A/MisTra/Part per gli errori riguardanti il trasferimento parziale o impreciso del significato (90

errori) e A/ForLa per gli errori relativi all'uso non richiesto della lingua inglese o all'impiego di parole non esistenti nella lingua di arrivo (65 errori). Di seguito vengono riportati due esempi per quest'ultima categoria.

Esempio n. 49 (Prog2_Testo5 – A/ForLa)

TO: Einen Text sprachlich verfremden heißt [...]

TA: Stranizzare linguisticamente un testo significa [...]

PE: Straniare linguisticamente un testo significa [...]

Nell'esempio n. 49, la TA propone il verbo “stranizzare”, lessema che, sebbene sia morfologicamente corretto e attestato in alcune varietà regionali siciliane, non rientra nell'italiano standard né nel linguaggio scientifico⁶². La correzione apportata in fase di *post-editing* propone il verbo “straniare”, forma consolidata e appropriata al linguaggio e all'ambito trattato. Questo caso mette in luce come la traduzione automatica possa generare forme morfologicamente corrette e attestate in varietà non standard, ma inadeguate al contesto del linguaggio scientifico.

Esempio n. 50 (Prog2_Testo3 – A/ForLa)

TO: Der Auswertung von Dieter Huber in “Mögliche Welten für die Übersetzungswissenschaft – Überlegungen zur Paradigmengeschichte der Translation” zufolge [...]

TA: Secondo l'analisi di Dieter Huber in “Possible Worlds for Translation Science - Reflections on the Paradigm History of Translation”, [...]

PE: Secondo l'analisi di Dieter Huber in “Mögliche Welten für die Übersetzungswissenschaft – Überlegungen zur Paradigmengeschichte der Translation”, [...]

Nell'esempio n. 50, la TA traduce il titolo dell'opera in inglese, sebbene il titolo originale sia in tedesco. Pertanto, in fase di *post-editing*, il titolo è stato riportato nella lingua originale. In questo caso, il sistema di TA introduce un uso non richiesto della lingua inglese, verosimilmente dovuto all'influenza dei corpora di addestramento, nei quali l'inglese occupa una posizione predominante. Tuttavia, nel contesto accademico e traduttologico, è uso conservare i titoli di pubblicazioni scientifiche nella lingua originale per garantire la correttezza filologica e la riconoscibilità della fonte.

Proseguendo nell'analisi degli errori, si individuano 61 occorrenze per la categoria A/MisOrd/WO. Questa riguarda le variazioni di significato dovute all'ordine delle parole che

⁶² Per maggiori informazioni si veda <https://accademiadellacrusca.it/it/consulenza/stranizzare--stranizzarsi/994> (ultima consultazione: 20/12/2025)

influenzano quindi la corretta trasmissione del messaggio di partenza nel testo di arrivo. Di seguito viene riportato un esempio.

Esempio n. 51 (Prog2_Testo5 – A/MisOrd/WO)

TO: Zentral für Fremdheitserfahrung in literarischer Übersetzung sind daher auch für Marek die von Huntemann und Rühling definierten extrinsischen Phänomene [...].

TA: I fenomeni estrinseci definiti da Huntemann e Rühling sono quindi centrali anche nell'esperienza dell'estraneità nella traduzione letteraria per Marek [...].

PE: I fenomeni estrinseci definiti da Huntemann e Rühling sono quindi centrali nell'esperienza dell'estraneità nella traduzione letteraria anche per Marek [...].

L'esempio n. 51 mostra come la variazione nell'ordine delle parole possa incidere sulla coerenza argomentativa e sulla corretta trasmissione delle relazioni concettuali nel linguaggio scientifico. L'intervento di *post-editing* focalizzato sullo spostamento di posizione degli elementi ripristina l'ordine sintattico e logico del testo di partenza, rendendo più chiara la relazione tra i concetti espressi.

Un'ulteriore categoria che quantitativamente non registra un numero consistente di errori è A/MisExpl (50 errori), che segnala di casi di inadeguatezza dei contenuti per il destinatario del testo di arrivo. L'esempio n. 53 ne riporta un caso.

Esempio n. 53 (Prog2_Testo7 – A/MisExpl)

TO: [...] durch die Aufschrift "Halt!" auf dem Verkehrszeichen-Piktogramm

TA: [...] dalla scritta "Stop!" sul pittogramma del segnale stradale

PE: [...] dalla scritta "Halt!" (Stop) sul pittogramma del segnale stradale

In questo esempio, la macchina traduce il segnale stradale tedesco "Halt!" con il corrispondente italiano "Stop!", scegliendo un equivalente funzionale. Tuttavia, nel testo di partenza il segnale è presente in una figura con la scritta originale "Halt!", quindi la traduzione automatica introduce un'informazione non conforme al contenuto reale della figura. In fase di *post-editing*, viene reintrodotta la scritta originale "Halt!" seguita dalla traduzione esplicativa tra parentesi, garantendo fedeltà al testo e comprensibilità per il lettore del testo di arrivo.

Una categoria che registra un numero contenuto di errori è F/De/Layo con 47 occorrenze. Si tratta di errori relativi alla strutturazione del testo e dei paragrafi che forniscono informazioni sulla gestione della segmentazione e della formattazione dei paragrafi da parte della macchina (cfr. esempio n. 52).

Esempio n. 52 (Prog2_Testo9 – F/De/Layo)

TO: (vgl. auch Stollhans 2012: 614).

TA: (cfr. anche
Stollhans 2012: 614).

PE: (cfr. anche Stollhans 2012: 614).

In questo caso, si nota che il sistema di TA introduce un'interruzione di riga non presente nel segmento di partenza, alterando la linearità della citazione bibliografica. In fase di *post-editing*, la formattazione viene corretta, riportando il riferimento su un'unica linea, come nel testo originale.

Tra le categorie di errori più rilevanti per la traduzione del linguaggio scientifico e del genere testuale dell'articolo scientifico, spicca la categoria F/Coh/Cit che conta 28 occorrenze e riguarda la strutturazione incoerente dei riferimenti bibliografici.

Una categoria di errori con risultati contenuti è A/MisTra/Gend (44 errori). Tale categoria segnala gli errori relativi alla mancata applicazione del linguaggio inclusivo (cfr. esempio n. 27 in 6.2.1), che in contesti accademici e scientifici è ormai uno standard consolidato. 28 occorrenze del totale si verificano nel testo n. 6 (cfr. Tabella 11 – allegato 3), da cui viene estratto l'esempio seguente che mostra l'omissione della forma femminile nel segmento di arrivo:

Esempio n. 54 (Prog2_Testo6 – A/MisTra/Gend)

TO: Im Fokus stehen Übersetzer*innen [...]

TA: L'attenzione è rivolta ai traduttori [...]

PE: L'attenzione è rivolta alle traduttrici e ai traduttori [...]

Altre categorie di errori a bassa incidenza sono: A/MisTra/WroPr con 41 errori (errori di preposizioni, sintagmi preposizionali o congiunzioni); A/MisTra/POS (errore di categoria lessicale) e F/Coh/CoRef (errori di mancata corrispondenza di referente in frasi diverse) con 39 errori; A/Mech/CharFor con 35 errori (errori di trasferimento della formattazione); A/Add con 32 errori (segmenti aggiunti); A/Oth con 20 errori e A/MWE/Komp con 17 errori (errori di traduzione di aggettivi composti). Anche in questo caso si tratta di errori di *Accuracy* e di *Fluency* che compromettono il significato generale del testo e la fedeltà al testo di partenza.

La categoria A/MWE/Komp è particolarmente interessante perché gli aggettivi composti costituiscono una caratteristica tipica del linguaggio scientifico tedesco, in cui la densità informativa è spesso concentrata in strutture morfologicamente complesse. Gli errori della categoria A/MWE/Komp derivano dunque da una traduzione letterale che non tiene conto di queste differenze strutturali, come illustrano gli esempi seguenti.

Esempio n. 55 (Prog2_Testo1 – A/MWE/Komp)

TO: soziologisch-hermeneutische Ansätze

TA: approcci sociologico-ermeneutici

PE: approcci ermeneutico-sociologici

Nel composto “soziologisch-hermeneutisch”, il secondo elemento (“hermeneutisch”) costituisce il determinato, mentre “soziologisch” funge da determinante e ne precisa l’ambito teorico: si tratta quindi di approcci ermeneutici che rientrano nel campo della sociologia. La traduzione automatica (“approcci sociologico-ermeneutici”) inverte il rapporto semantico nel *Determinativkompositum*, suggerendo invece che gli approcci siano di natura sociologica con un orientamento ermeneutico e spostando dunque il centro semantico del composto. L’intervento di *post-editing* ristabilisce l’ordine interpretativo corretto, riflettendo la struttura concettuale del termine tedesco e preservando la gerarchia tra determinante e determinato nella lingua italiana.

Esempio n. 56 (testo n. 8) (Prog2_Testo8 – A/MWE/Komp)

TO: die sprachlich-kulturelle Anpassung

TA: l’adattamento linguistico e culturale

PE: l’adattamento linguistico-culturale

Nell’esempio n. 56, il composto “sprachlich-kulturell” rappresenta un’unica unità concettuale, tipica del linguaggio scientifico tedesco, in cui due aggettivi coordinati formano un composto coordinato, cioè un *Kopulativkompositum*. Nella proposta della TA i due aggettivi vengono separati, riducendo la coesione del sintagma. Nella versione post-editata è stato ripristinato l’aggettivo composto (“linguistico-culturale”), conforme anche all’uso del linguaggio scientifico italiano.

Per quanto riguarda la categoria A/Oth, questa segnala errori che non possono essere ricondotti alle altre categorie di errori già esistenti. In questo progetto di traduzione, tale categoria è stata principalmente impiegata per classificare le traduzioni errate di citazioni in tedesco per le quali esistono versioni ufficiali in italiano, non presentate dal sistema di TA. L’esempio n. 57 mostra nel testo di partenza una citazione tratta dal romanzo “Das Böse Mädchen” di Mario Vargas Llosa. In fase di *post-editing*, la proposta del sistema di TA è stata sostituita con la frase corrispettiva pubblicata nella versione italiana del romanzo, “Avventure della ragazza cattiva” (Vargas Llosa 2006).

Esempio n. 57 (Prog2_Testo6 – A/Oth)

TO: Es war komplizierter als die Reden und Redebeiträge zu übersetzen, an die ich in meiner Arbeit gewöhnt war.

TA: È stato più complicato che tradurre i discorsi e gli interventi a cui ero abituato nel mio lavoro.

PE: È stata un'impresa più difficile che tradurre i discorsi e gli interventi ai quali ero abituato nel mio lavoro.

Infine, le categorie meno frequenti includono F/Sty/Idiom (15 occorrenze), A/Mech/Capit (13 occorrenze), A/MisOrd/MW (11 occorrenze) A/MWE/Ell (6 occorrenze), che indicano problematiche relative alla traduzione di espressioni idiomatiche (F/Sty/Idiom), all'attribuzione semantica di elementi (A/MisOrd/MW), al trasferimento della lettera maiuscola (A/Mech/Capit) o alla traduzione di parole composte (A/MWE/Ell) risultano quasi trascurabili.

Nel complesso, gli errori strettamente correlati alla traduzione del linguaggio scientifico e dei suoi aspetti caratterizzanti ammontano a 1.262, circa il 18,3% del totale. Si tratta di categorie che riflettono direttamente i tratti distintivi propri della scrittura scientifica, come la gestione della terminologia specialistica e l'uso delle nominalizzazioni o delle strutture impersonali, e che richiedono quindi un intervento umano mirato per garantire un risultato coerente con le convenzioni del linguaggio scientifico. Tuttavia, anche altre tipologie di errore, come quelle relative all'*Accuracy*, alla *Fluency* e alla *Grammar*, incidono sulla resa complessiva del linguaggio scientifico, influenzandone la chiarezza espositiva, la precisione concettuale e la correttezza formale.

Nel loro insieme, i risultati offrono spunti significativi per riflettere sull'effettiva affidabilità della TA applicata al linguaggio scientifico tedesco in italiano: pur producendo esiti utili, il sistema non raggiunge ancora standard pienamente pubblicabili né garantisce costanza nell'accuratezza semantica o nella naturalezza linguistica. Il contributo del *post-editor* risulta pertanto determinante.

6.3.2 Analisi dell'*effort* temporale e tecnico

La durata totale delle sessioni per il progetto di traduzione *trans-kom* corrisponde a 115:45:51, con una durata media per testo di 12:51:45. Le sessioni di *post-editing* superiori a 01:00:00 circa sono state intervallate da brevi interruzioni ogni 01:30:00/01:50:00, non incluse nel computo complessivo. Le indicazioni temporali relative a ciascun testo sono riportate in dettaglio nell'allegato 3.

Dall'analisi del report di Post-Edit Compare emerge che il progetto ha coinvolto nove testi, per un totale di 7.304 segmenti e 127.112 parole. Di questi, la quasi totalità (7.048 segmenti) è stata prodotta dal sistema di TA DeepL, mentre soltanto 256 segmenti sono stati inseriti manualmente poiché il sistema di TA non ha proposto una traduzione. È importante precisare che la memoria di traduzione, creata appositamente per il progetto, era inizialmente vuota e si è popolata

progressivamente durante la fase di *post-editing*. Tuttavia, questo aggiornamento non influisce sui risultati del report: Post-Edit Compare opera, infatti, a livello di progetto e confronta esclusivamente la versione iniziale salvata prima dell'avvio del *post-editing* (costituita unicamente dall'output della TA) con la versione finale post-editata. Di conseguenza, anche se la memoria si è arricchita nel corso del lavoro, essa non interviene nel confronto e non può generare *matches* nei dati analizzati. Ciò conferma che il testo di partenza preso in esame da Post-Edit Compare corrisponde integralmente all'output prodotto dalla TA.

La Tabella 7 illustra le percentuali complessive rilevate nelle diverse fasce di PEMp, riportando per ciascuna di esse il numero associato di segmenti e di parole e consentendo di ottenere una visione completa della distribuzione delle tipologie di interventi.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	2.572	26.252	20,65%
95-99%	790	20.492	16,12%
85-94%	1.210	32.054	25,22%
75-84%	598	13.912	10,94%
50-74%	540	10.112	7,96%
New (0-49%)	1594	24290	19,11%

Tabella 7: Distribuzione delle fasce di PEMp per progetto n. 2

L'analisi della distribuzione mostra una certa polarizzazione delle fasce di interventi: il 20,65% delle parole rientra nella fascia 100% (2.572 segmenti, 26.252 parole), segnalando che una parte considerevole del testo (un quinto) è stata accettata senza modifiche. Nella Figura 32 viene riportato un segmento appartenente a questa fascia di PEMp. Si può osservare nella parte in alto della figura il segmento di partenza per il quale il sistema di TA propone una traduzione che è accettata senza modifiche. Per questo motivo la casella del *post-editing* appare vuota nel report. La colonna PEMp riporta la percentuale 100% a indicare che non è stata effettuata alcuna modifica, informazione confermata dal dato dell'*edit-distance* pari a 0.

Testo di partenza
Renn unterscheidet dabei mindestens zwei Typen kultureller Differenzen.

Testo di partenza	Post-editing	PEMp
Renn distingue almeno due tipi di differenze culturali.		100% Edit-Dist.: 0 Max chars: 55

Figura 32: Esempio fascia di PEMp 100% dal testo n. 1 progetto n. 2

Il 16,12% delle parole rientra nella fascia 95-99% (790 segmenti, 20.492 parole), che indica le correzioni minime. Tuttavia, la quota più consistente si colloca nella fascia 85-94%, con il 25,22% (1.210 segmenti, 32.054 parole), suggerendo che un quarto del totale delle parole è stato sottoposto a correzioni di media entità. Nella Figura 33 viene fornito un esempio per questa fascia di PEMp, in cui si possono osservare, nella colonna del confronto, le modifiche apportate rispetto alla proposta del sistema di TA. La percentuale di PEMp in questo caso ammonta a 86,36%.

Testo di partenza
So erhoben die semantisch-linguistischen Äquivalenzmodelle, die seit den 60er-Jahren des 20. Jahrhunderts entstanden sind, den Anspruch der Objektivierbarkeit von Sinn, weswegen sie für die Klärung der Frage der Bewertbarkeit von Translaten gut geeignet zu sein schienen.

Testo di partenza	Post-editing	Confronto	PEMp
I modelli di equivalenza semantico-linguistica emersi a partire dagli anni Sessanta pretendono di oggettivare il significato, e per questo motivo sembrano adatti a chiarire la questione della valutabilità dei tradotti.	I modelli di equivalenza linguistico-semantica, emersi a partire dagli anni Sessanta, pretendono di oggettivare il senso, e per questo motivo sembrano adatti a chiarire la questione della valutabilità dei testi tradotti.	I modelli di equivalenza ~~semantico-linguistica~~ <u>linguistico-semantica</u> , emersi a partire dagli anni Sessanta, pretendono di oggettivare il ~~significato~~ <u>senso</u> , e per questo motivo sembrano adatti a chiarire la questione della valutabilità dei <u>testi</u> tradotti.	86.36% Edit-Dist.: 30 Max chars: 220

Figura 33: Esempio fascia di PEMp 85-94% dal testo n. 1 del progetto n. 2

La fascia 75-84% registra il 10,94% (598 segmenti, 13.912 parole), mentre la fascia 50-74% copre il 7,96% (540 segmenti, 10.112 parole), suggerendo che le modifiche di entità media e medio-alta hanno una diffusione limitata. Più significativo è il peso della fascia *New* (0-49%), pari al 19,11% (1.594 segmenti, 24.290 parole), che attesta la necessità di riscrivere, in larga misura o integralmente, quasi un quinto degli output del progetto elaborati dalla TA. La Figura 34 mostra un caso in cui il sistema di TA non ha proposto alcuna traduzione e pertanto, in fase di *post-editing*, è stato necessario aggiungerla manualmente.

Testo di partenza	TA	Post-editing	PEMp
<symbol/> <symbol/> <cf>Fehlen von Konnotationen, von affektiven und wertenden Wortkomponenten, Wörtern und Wendungen; </cf>		<symbol/> <symbol/> <cf>Assenza di connotazioni, di componenti lessicali, parole ed espressioni affettive e valutative; </cf>	0% Edit-Dist.: 0 Max chars: 100

Figura 34: Esempio categoria di PEMp New dal testo n. 4 del progetto n. 2

La colonna indicante la traduzione automatica è vuota, mentre la traduzione presente nella colonna del *post-editing* è scritta in blu ed è evidenziata in giallo, a indicare che è stata inserita interamente dal *post-editor*. Ciò è confermato dai dati di PEMp e *edit-distance*, pari a 0.

Nel complesso, l'analisi mette in luce un quadro marcatamente disomogeneo. Il dato della fascia 100% (20,65%) è da considerarsi significativo, in quanto consente di dedurre che il sistema di TA ha raggiunto una qualità pubblicabile già per un quinto del totale delle parole. Tuttavia, per la restante parte, l'intervento del *post-editor* si è rivelato indispensabile e ha comportato modifiche di varia entità. In particolare, quasi un quinto del testo della TA è stato rielaborato in modo integrale o pressoché integrale, contrapponendosi al valore della fascia 100% e ridimensionandone così la connotazione positiva. Circa un quarto del testo, invece, ha richiesto solo correzioni lievi, mentre i restanti hanno subito interventi di intensità variabile, da minima a medio-alta. Ne risulta quindi una qualità complessivamente irregolare, che solleva diversi dubbi sull'effettiva efficienza del sistema di TA in progetti di traduzione di questo tipo. Tali perplessità trovano ulteriore conferma nell'analisi qualitativa, che mette in evidenza errori sistematici di natura terminologica, sintattica e stilistica, indicativi di limiti strutturali del sistema di TA nel trattamento di testi scientifici.

Il risultato della TA può essere considerato un punto di partenza promettente; tuttavia, è necessario ricordare che la qualità varia sensibilmente da un segmento all'altro. Di conseguenza, il *post-editor* deve modulare costantemente il proprio intervento in funzione della complessità delle revisioni richieste. In termini di *effort*, il *post-editing* non si limita a semplici correzioni, ma implica spesso rielaborazioni estese, riducendo così i vantaggi attesi dall'automazione.

6.4 Progetto n. 3: *Interculture journal*

Si procede in questa sezione con l'analisi complessiva dei testi che costituiscono il progetto di traduzione *Interculture journal*. Il numero di *tokens* complessivo del progetto ammonta a 76.232.

6.4.1 Analisi degli errori⁶³

⁶³ Le analisi dei singoli testi sono state riportate nell'allegato 4.

Categoria di errore	Somma annotazioni
F/Coh/InPu	1667
A/MisTra/SemUn	849
F/Sty/Disfl	833
F/Lex/LexCho	649
A/Omis	226
G/Orth/Capit	224
A/MisTra/WSD	179
A/Untra	168
A/MisTra/Part	162
A/DNT	152
A/MisTra/Gend	143
G/Orth/Punct	137
F/SLP/Nomin	120
F/De/Layo	101
F/Coh/Cit	89
F/Sty/Reg	89
A/MisTra/POS	83
F/De/CharFor	82
G/Orth/WhiSpa	79
A/Mech/Punct	77
A/MisOrd/WO	76
A/Add	64
F/Lex/Fue	63
F/SLP/ForImp	62
A/MWE/Komp	60
F/Coh/CoRef	58
A/SouEr	57
G/Orth/Spel	51
A/Termi/DNT	49
A/MisTra/WroPr	45
A/Termi/SemUn	44
G/WF/Agree	42
A/MisOrd/MW	34
A/ForLa	25
F/Sty/Idiom	20
A/Mech/CharFor	20
A/MisExpl	16
A/MWE/Ell	12
G/WF/VF	11
A/Oth	10
A/Mech/Capit	5

Tabella 8: Analisi quantitativa degli errori per progetto n. 3

Nel progetto di traduzione si rilevano complessivamente 6.933 errori, suddivisi in 3.833 casi di *Fluency*, 2.556 di *Accuracy* e 544 di *Grammar*. La distribuzione evidenzia come la

macrocategoria più frequente sia la *Fluency*, in particolare per l'alta incidenza degli errori di F/Coh/InPu (1.667 occorrenze), che segnala l'uso incoerente di segni ortografici, quali l'apostrofo e le virgolette (cfr. esempio n. 1 in 6.2.1 ed esempio n. 31 in 6.3.1). Come evidenziano le analisi quantitative dei dati riportate nell'allegato 4, questa categoria è quella maggiormente rappresentata in tutti testi. Si ribadisce, in linea con quanto osservato nelle analisi complessive precedenti, che tali errori non dipendono dalla specificità del linguaggio scientifico, rappresentano tuttavia problematiche tecniche tipiche della TA in generale.

Seguono per frequenza 849 errori di A/MisTra/SemUn e 833 di F/Sty/Disfl. Si tratta, nel primo caso, di errori di corrispondenza semantica tra il segmento di partenza e quello tradotto e, nel secondo caso, di errori afferenti alla scorrevolezza della lingua di arrivo. Gli errori di queste due categorie incidono pertanto sulla leggibilità e chiarezza, aspetti essenziali nei testi scientifici. Queste categorie di errori sono tra le maggiormente rappresentate nei testi analizzati: come si può osservare nelle tabelle quantitative degli errori presenti nell'allegato 4, occupano posizioni che variano tra il secondo e il quarto posto. Di seguito sono presentati quattro esempi di *post-editing*: i numeri 61 e 62 riguardano la categoria A/MisTra/SemUn, 63 e 64 la categoria F/Sty/Disfl.

Esempio n. 58 (Prog3_Testo1 – A/MisTra/SemUn)

TO: [...] dass in diesem Theorieangebot ein unmittelbar anschlussfähiges Konzept für die Analyse des Zusammenhangs von Kultur, Denken und Handeln gesehen werden kann,

TA: [...] che questa proposta teorica può essere vista come un concetto direttamente collegabile per analizzare la connessione tra cultura, pensiero e azione

PE: [...] che in questa proposta teorica può essere visto un concetto direttamente collegabile all'analisi della connessione tra cultura, pensiero e azione

Esempio n. 59 (Prog3_Testo4 – A/MisTra/SemUn)

TO: diese holzschnittartig darstellen

TA: le presentano in maniera silvestre

PE: le presentano in maniera rudimentale

Negli esempi n. 58 e n. 59, il sistema di TA presenta proposte di traduzione che alterano il senso della frase di partenza. In fase di *post-editing* si è intervenuto con correzioni mirate, ripristinando il senso originale.

Esempio n. 60 (Prog3_Testo2 – F/Sty/Disfl)

TO: [...], die ihm ganz konkrete Hinweise dafür liefert, wie er trotz komplexer Verhältnisse handlungsfähig bleiben kann.

TA: [...], che gli fornisce indicazioni molto concrete su come può rimanere capace di agire nonostante le circostanze complesse.

PE: [...], che gli fornisce indicazioni molto concrete su come può mantenere la sua capacità di agire nonostante le circostanze complesse.

L'esempio n. 60 invece riguarda la naturalezza sintattica del testo di arrivo: la versione post-editata migliora la coesione e la scorrevolezza, evitando costruzioni rigide o calchi che ostacolano la leggibilità.

Risulta rilevante anche la categoria F/Lex/LexCho (647 occorrenze), che segnala l'uso di espressioni non conformi all'ambito specialistico trattato e al genere testuale dell'articolo scientifico. Gli esempi n. 61 e n. 62 ne mostrano due casi. Come già notato per le analisi complessive precedenti, anche in questo caso tale categoria rientra tra quelle che registrano più risultati, collocandosi tra il secondo e il quinto posto nelle tabelle quantitative degli errori illustrate nell'allegato 4.

Esempio n. 61 (Prog3_Testo3 – F/Lex/LexCho)

TO: Simons beschreibt hier in aller Kürze einen Tatbestand, der aktuell in unterschiedlichen Disziplinen [...] zunehmend diskutiert wird.

TA: Simons describe brevemente un fatto che attualmente viene sempre più discusso in varie discipline [...]

PE: Simons describe brevemente in questo passo una circostanza che attualmente viene sempre più discussa in varie discipline [...]

Esempio n. 62 (Prog3_Testo6 – F/Lex/LexCho)

TO: [...] da die Geschichte eines universitären Fachs häufig vergessen wird [...]

TA: [...] dal momento che la storia di una materia universitaria è spesso dimenticata [...]

PE: [...] dal momento che la storia di una disciplina universitaria è spesso dimenticata [...]

In entrambi i due esempi di F/Lex/LexCho, le proposte di *post-editing* sono volte ad affinare la scelta lessicale per garantire un livello di specializzazione linguistica adeguata al genere testuale e all'ambito specialistico presi in esame. Nel primo caso, inoltre, il *post-editing* introduce l'espressione "in questo passo", che ripristina l'ancoraggio deittico espresso da "hier" nel testo originale. Tale elemento, omesso nella traduzione automatica e quindi interessato da un errore di A/Omis, contribuisce a mantenere la coerenza testuale e il corretto riferimento discorsivo all'interno del testo.

Per quanto concerne la macrocategoria *Grammar*, essa riguarda aspetti di correttezza della lingua di arrivo. Si rilevano 224 errori relativi all'impiego della lettera maiuscola (G/Orth/Capit), 137 errori di punteggiatura (G/Orth/Punct), 79 errori di spaziatura (G/Orth/WhiSpa), 51 errori di scrittura di parole (G/Orth/Spel), 42 errori riguardanti l'accordo tra parole (G/WF/Agree) e 11 errori riferibili all'uso grammaticalmente scorretto di tempi o forme verbali (G/WF/VF). Pur non incidendo sul linguaggio scientifico in senso stretto, questi errori assumono rilevanza in quanto condizionano la qualità del linguaggio e la pubblicabilità del testo. La Tabella 3 nell'allegato 4 mostra il ruolo chiave assunto dalla categoria G/Orth/Capit nel testo n. 2, di cui di seguito viene riportato un esempio in cui la lettera maiuscola è stata usata impropriamente.

Esempio n. 63 (Prog3_Testo2 – G/Orth/Capit)

TO: Jedoch zeigt die Forschung: Menschliche Interaktionen, [...]

TA: Tuttavia, la ricerca mostra che Le interazioni umane [...]

PE: Tuttavia, la ricerca mostra che le interazioni umane [...]

Le categorie di errori G/WF/Agree e G/WF/VF possono incidere più direttamente sulla comprensione del testo. Di seguito si riportano due casi esemplificativi.

Esempio n. 64 (Prog3_Testo7 – G/Orth/Agree)

TO: Schlagwörter: [...], pädagogische Handlungsstrategien, bedeutsame Situationen

TA: Parole chiave: [...], pedagogico strategie di azione, significativo situazioni

PE: Parole chiave: [...], strategie pedagogiche di azione, situazioni significative

L'esempio n. 64 mostra un errore di accordo morfosintattico tra aggettivi e sostantivi (G/Orth/Agree). Nei testi scientifici, gli elenchi di parole chiave rappresentano una sfida per i sistemi di TA neurale. Questi modelli sono progettati per tradurre sequenze contestualizzate, e pertanto l'assenza di una struttura sintattica completa e di un contesto sufficiente può generare errori di concordanza, traduzioni non scorrevoli o lessemi non correttamente adattati alla morfosintassi della lingua di arrivo. In questo caso il modello ha tradotto parola per parola, replicando la struttura superficiale del testo originale senza applicare le regole di concordanza e flessione necessarie in italiano. Inoltre, l'errore si combina con un caso di F/Sty/Disfl, relativo alla scorrevolezza del testo: nella proposta di *post-editing* gli aggettivi sono stati posposti ai sostantivi nelle versioni post-editate, per rispettare la naturale collocazione sintattica in italiano e migliorare la leggibilità.

Esempio n. 65 (Prog3_Testo6 – G/WF/VF)

TO: Inwieweit Thomas' Übertragung dieser Prämisse auf die aktuellen Lebensumstände gerechtfertigt ist, wird im Folgenden einer kritischen Prüfung unterzogen.

TA: Di seguito verrà esaminata criticamente la misura in cui il trasferimento di questa premessa da parte di Thomas alle condizioni di vita attuali è giustificato.

PE: La misura in cui sia giustificato il trasferimento da parte di Thomas di questa premessa alle condizioni di vita attuali verrà esaminata criticamente di seguito.

L'esempio n. 65 per la categoria G/WF/VF riguarda invece l'uso improprio dei tempi e modi verbali. La sostituzione con il congiuntivo "sia giustificato" contribuisce a restituire il valore potenziale e argomentativo dell'enunciato, coerente con la funzione di analisi critica del testo scientifico. Il segmento è inoltre interessato da un errore di F/Sty/Disfl, riguardante la scorrevolezza della lingua di arrivo, che ha richiesto un intervento di riformulazione per migliorare la naturalezza e la leggibilità complessiva della frase.

Un'attenzione particolare meritano alcuni errori di *Accuracy* che, pur non riguardando direttamente le caratteristiche del linguaggio scientifico, incidono per il numero consistente di occorrenze sulla chiarezza e sulla fedeltà al testo di partenza. Tra questi si notano: errori di omissioni di segmenti del testo di partenza nel testo di arrivo (A/Omis, 226 errori) ed errori di scelta semantica in riferimento a parole polisemiche (A/MisTra/WSD, 176 errori). La categoria A/Omis riveste un ruolo degno di nota nel testo n.1 (cfr. Tabella 1 – allegato 4), dove si sono verificate numerose omissioni di parole e frasi. Di seguito ne viene riportato un caso che crea problemi di precisa ricezione del testo.

Esempio n. 66 (Prog3_Testo1 – A/Omis)

TO: Diesseits bzw. jenseits von Kultur lagen bei diesem Coaching die Themen Teambildung und Teamprozesse [...]

TA: Questo coaching si è concentrato sui temi del team building e dei processi di gruppo [...]

PE: In termini culturali ed extra-culturali, questo coaching si è concentrato sui temi del team building e dei processi di gruppo [...]

Interessanti sono anche le categorie A/Untra con 168 errori e A/MisTra/Part con 162 errori. Come emerge dalla Tabella 17 nell'allegato 4, queste categorie di errori risultano particolarmente significative nel testo n. 9. La categoria A/Untra segnala i casi in cui il sistema di TA non ha fornito una traduzione (cfr. esempio n. 67).

Esempio n. 67 (Prog3_Testo9 – A/Untra)

TO: Abschließend werden Schlussfolgerungen für Forschung und Praxis abgeleitet.

TA: Abschließend werden Schlussfolgerungen für Forschung und Praxis abgeleitet.

PE: Infine, vengono tratte conclusioni utili sia per la ricerca sia per la pratica.

Nel segmento preso in esame, il sistema di TA non ha prodotto alcuna traduzione, lasciando il testo identico all'originale. Tale fenomeno rientra nel più generale problema dell'*under-translation*, già citato in riferimento all'esempio n. 13 del progetto n. 1. Nel caso in esame, l'intervento del *post-editor* è stato necessario per inserire manualmente la traduzione corretta, restituendo completezza e coerenza al testo.

La categoria A/MisTra/Part indica gli errori di trasferimento semantico parziale o impreciso. Nel caso specifico illustrato nell'esempio n. 68, tratto dal testo n. 9, si tratta dell'uso della forma maschile in sostituzione della forma femminile richiesta.

Esempio n. 68 (Prog3_Testo9 – A/MisTra/Part)

TO: Eine andere Probandin dieser Gruppe ergänzt: [...]

TA: Un altro intervistato di questo gruppo aggiunge: [...]

PE: Un'altra intervistata di questo gruppo aggiunge: [...]

Questo fenomeno può essere spiegato alla luce del *gender bias* nei sistemi di TA neurale. Come osservano Stanovsky et al. (2019), nelle lingue che prevedono la declinazione di genere, il modello tende a privilegiare le forme di genere presenti con maggiore frequenza nei dati di addestramento, generando traduzioni che non rispettano il genere corretto del testo originale.

Si registra un numero considerevole di annotazioni anche per la categoria A/DNT (152 errori). Si tratta di un'ulteriore categoria di *Accuracy* che influenza il contenuto e, di conseguenza, l'accuratezza complessiva del testo. In particolare, questi errori segnalano i casi di traduzione non necessaria. Tale categoria è particolarmente significativa nel testo n. 3, in cui sono presenti numerose citazioni in lingua inglese. In fase di *post-editing* si è deciso di non tradurle per rispettare la scelta dell'autore di inserirle in lingua originale. L'esempio n. 69 ne riporta un caso.

Esempio n. 69 (Prog3_Testo3 – A/DNT)

TO: These routines are mere communicative conventions and they do not necessarily need to be connected to culture.

TA: Queste routine sono mere comunicativo convenzioni e sono non non necessariamente necessario per essere collegato alla cultura.

PE: These routines are mere communicative conventions and they do not necessarily need to be connected to culture.

Nel caso sopra citato, la proposta di TA mostra diversi problemi riconducibili al funzionamento dei sistemi neurali. In primo luogo, il modello non ha riconosciuto che il testo originale era intenzionalmente in lingua inglese, trattandolo come un segmento da tradurre. Questo si verifica poiché la TA neurale agisce su sequenze di *token* e considera l'intero segmento traducibile, senza applicare strategie specifiche per preservare elementi che, nell'intenzione di chi scrive il testo di partenza, non richiedono traduzione. In secondo luogo, anche la traduzione in italiano risulta problematica: viene effettuata in modo letterale, parola per parola, senza rispettare le strutture sintattiche e la correttezza morfosintattica del testo. Ne derivano espressioni scorrette, come errori di concordanza e di flessione morfosintattica tra gli aggettivi e i sostantivi (“mere comunicativo convenzioni”) e alcune forme verbali sono ripetute o inappropriate (“non non necessariamente necessario”). Tali problematiche evidenziano il ruolo essenziale del *post-editor*, necessario per ripristinare la correttezza linguistica e garantire il rispetto della scelta autoriale di mantenere le citazioni in lingua originale.

Inoltre, sono stati annotati 143 errori per la categoria A/MisTra/Gend, che segnala le problematiche nell'uso del linguaggio inclusivo. Questa categoria risulta particolarmente rilevante per i testi n. 7 e n. 9, dove si registrano rispettivamente 60 e 67 occorrenze (cfr. tabelle 13 e 17 – allegato 4). Di seguito viene riportato un esempio.

Esempio n. 70 (Prog3_Testo9 – A/MisTra/Gend)

TO: [...] zwischen den Schülerinnen und Schülern [...]

TA: [...] tra_gli studenti [...]

PE: [...] tra le studentesse e gli studenti [...]

In riferimento ai fenomeni linguistici caratterizzanti il linguaggio scientifico, emergono 120 errori di F/SLP/Nomin, riguardanti la mancata realizzazione di nominalizzazioni (cfr. esempio n. 71) e 62 errori di F/SLP/ForImp, connessi all'errato uso di forme personali in sostituzione delle più adeguate impersonali (cfr. esempio n. 72).

Esempio n. 71 (Prog3_Testo2 – F/SLP/Nomin)

TO: Durch das Bestimmen der Elemente und [...]

TA: Identificando gli elementi e [...]

PE: Attraverso l'identificazione degli elementi e [...]

Nell'esempio n. 71, la TA tende a rendere la struttura nominale tedesca (“Durch das Bestimmen der Elemente”) con una forma verbale (“Identificando gli elementi”), più naturale nel linguaggio comune ma meno adeguata al linguaggio scientifico, poiché la costruzione verbale introduce una

prospettiva agentiva e riduce l'astrattezza e la compattezza tipica dei sintagmi nominali. L'intervento di *post-editing* ristabilisce la costruzione nominale, più aderente alle convenzioni stilistiche del testo scientifico.

Esempio n. 72 (Prog3_Testo3 – F/SLP/ForImp)

TO: Im weitesten Sinne kann hier auch von [...] gesprochen werden.

TA: In senso lato, possiamo anche parlare di [...]

PE: In senso lato, si può anche parlare di [...]

Nell'esempio n. 72, la TA opta per una forma personale ("possiamo parlare"), mentre la versione post-editata ripristina la costruzione impersonale ("si può parlare"), più conforme alle convenzioni del discorso scientifico italiano, che mira a un tono oggettivo e distaccato.

Per quanto attiene lo stile del linguaggio scientifico, si segnalano 88 errori di F/Sty/Reg, che indicano l'impiego di un registro linguistico non idoneo al genere testuale. L'esempio seguente mostra un caso in cui il sistema di TA ricorre a un'espressione colloquiale, non conforme al registro formale proprio del linguaggio scientifico.

Esempio n. 73 (Prog3_Testo4 – F/Sty/Reg)

TO: fallen zunächst eine Reihe von [...] ins Auge.

TA: inizialmente saltano all'occhio una serie di [...]

PE: inizialmente emergono una serie di [...]

In relazione al lessico del linguaggio scientifico, sono stati riscontrati 63 errori di F/Lex/Fue, riferibili alla traduzione errata delle *Fügungen*. Di seguito viene riportato un esempio:

Esempio n. 74 (Prog3_Testo8 – F/Lex/Fue)

TO: Im Folgenden soll exemplarisch dargestellt werden, wie [...]

TA: Ecco un esempio di come [...]

PE: Di seguito viene riportato a titolo esemplificativo come [...]

In questo caso, il sistema di TA rende la *Fügung* con la costruzione più vicina al linguaggio comune e informale. La versione post-editata, invece, privilegia una combinazione che rispetta le convenzioni del linguaggio scientifico italiano: utilizza una locuzione fissa ("Di seguito" per "Im Folgenden") adeguata al linguaggio scientifico e mantiene la densità informativa della *Fügung* originale. Questo tipo di errore è particolarmente rilevante nei testi scientifici, poiché una traduzione non conforme alle formule lessicali consolidate può compromettere sia la precisione concettuale sia la coerenza stilistica del testo.

Un ruolo di rilievo è assunto, inoltre, dalla categoria F/Coh/Cit, con 89 occorrenze, che segnala incoerenze nella strutturazione dei riferimenti bibliografici, elemento centrale nel linguaggio scientifico e nel genere testuale dell'articolo scientifico (cfr. esempio n. 75). Questa categoria è particolarmente rilevante nel testo n. 8, come si può evincere dalla Tabella 15 nell'allegato 4.

Esempio n. 75 (Prog3_Testo8 – F/Coh/Cit)

TO: (Fuhse & Mützel 2010, Stegbauer 2016).

TA: (Fuhse & Mützel 2010, Stegbauer 2016).

PE: (Fuhse & Mützel 2010, Stegbauer 2016).

Nella proposta della TA, viene mantenuto il simbolo “&”, tipico delle convenzioni anglosassoni. L'intervento di *post-editing*, invece, sostituisce tale simbolo con la congiunzione “e”, conformemente alle norme redazionali adottate da *Interculture journal*, che si basano sul sistema autore-data del *Chicago Manual of Style* (17^a ed.)⁶⁴. La modifica non altera la struttura della citazione, ma ne garantisce la coerenza linguistica e l'uniformità stilistica all'interno di un testo redatto in italiano.

Ulteriori categorie con valori intermedi comprendono: errori di strutturazione del testo e dei paragrafi (F/De/Layo con 101 occorrenze); errori di categoria lessicale (A/MisTra/POS con 83 occorrenze); errori di formattazione e convenzioni tipografiche del testo di arrivo (F/De/CharFor con 82 occorrenze). La categoria F/De/CharFor mostra nel testo n. 9 una frequenza più elevata rispetto agli altri casi analizzati (70 occorrenze), come si può evincere dalla Tabella 17 nell'allegato 4. L'esempio seguente (76) mostra un caso per questa categoria di errore, in cui per la parola interessata viene annotata anche una mancata corrispondenza semantica (A/MisTra/SemUn), in quanto la traduzione proposta comporta una perdita del riferimento specifico al sistema scolastico tedesco, per cui nel *post-editing* si è preferito inserire un prestito identico.

Esempio n. 76 (Prog3_Testo9 – F/De/CharFor)

TO: [...] Sichtweisen von Schülerinnen und Schülern der Gymnasialklasse [...]

TA: [...] le opinioni delle alunne e degli alunni della classe del ginnasio [...]

PE: [...] le opinioni delle alunne e degli alunni della classe del Gymnasium [...]

Proseguendo in ordine decrescente si individuano: errori di trasferimento della punteggiatura (A/Mech/Punct con 77 occorrenze); errori riguardanti la variazione di significato dovuta all'ordine

⁶⁴ Per maggiori informazioni sulle norme redazionali di “Interculture journal” si veda <https://journals.uni-jena.de/ij/libraryFiles/downloadPublic/1> (ultima consultazione: 20/12/2025)

delle parole (A/MisOrd/WO con 76 occorrenze); errori relativi a segmenti aggiunti nel testo di arrivo (A/Add con 64 occorrenze); errori di traduzione di aggettivi composti (A/MWE/Komp con 60 occorrenze). In questo caso si tratta di errori di *Accuracy* e di *Fluency* che incidono sul testo generale, sul suo significato e la sua scorrevolezza. Di seguito viene riportato un esempio di errore per la categoria A/MWE/Komp, che riguarda gli aggettivi composti la cui traduzione risulta letterale e quindi errata. Questa categoria registra una quota di occorrenze particolarmente significativa nel testo n. 1, come si può dedurre dalla Tabella 1 nell'allegato 4.

Esempio n. 77 (Prog3_Testo1 – A/MWE/Komp)

TO: [...] kann [...] symbolisch-interaktionistisch [...] verstanden werden.

TA: può essere intesa in termini [...] simbolico-interazionistici.

PE: può essere intesa in termini [...] interazionistico-simbolici.

Nel composto “symbolisch-interaktionistisch”, “interaktionistisch” costituisce il determinato, “symbolisch” funge da determinante. Come nell'esempio n. 56, la TA inverte il rapporto semantico, compromettendo il significato originario. L'intervento di *post-editing* è utile, pertanto, a ristabilire l'ordine corretto della struttura determinante-determinato, preservando la coerenza concettuale.

Ulteriori categorie con valori intermedi comprendono: errori di mancata corrispondenza di referente in frasi diverse (F/Coh/CoRef con 58 occorrenze); errori presenti nel testo di partenza (A/SouEr con 57 occorrenze); errori di preposizioni, sintagmi preposizionali o congiunzioni (A/MisTra/WroPr con 45 occorrenze). Anche in questo caso si tratta di errori che hanno un'influenza sulla correttezza semantica e sulla scorrevolezza della lingua di arrivo. L'esempio seguente mostra il caso per la categoria A/MisTra/WroPr.

Esempio n. 78 (Prog3_Testo1 – A/MisTra/WroPr)

TO: Herausforderung für das interkulturelle Wissen

TA: Sfida della conoscenza interculturale

PE: Sfida per la conoscenza interculturale

Nell'esempio n. 78, l'errore consiste nella resa della preposizione “für”: la TA ha impiegato la preposizione “della” piuttosto che “per la”, alterando così il rapporto semantico espresso nel testo di partenza.

Gli errori di A/SouEr sono principalmente dovuti a posizionamenti errati dei numeri delle note a piè di pagina ed è una problematica che si riscontra nel file traducibile generato da Trados Studio, ma non nel file originale (cfr. esempio n. 18 in 6.2.1). Altri casi riguardano l'aggiunta non necessaria o l'omissione di punteggiatura, come mostrato nell'esempio n. 79 per il punto fermo.

Esempio n. 79 (Prog3_Testo8 – A/SouEr)

TO: [...] die Welt ein stückweit zu verändern (Bergmann 2007:12)

TA: [...] a cambiare un po' il mondo (Bergmann 2007:12)

PE: [...] a cambiare un po' il mondo (Bergmann 2007:12).

Si riscontrano inoltre casi di trasferimento impreciso della formattazione dei caratteri dal file originale del testo di partenza al file del testo di partenza creato dal CAT tool. Infine, alcuni errori di A/SouEr sono dovuti a parole scritte in modo scorretto nel testo di partenza, compromettendo la possibilità per il sistema di TA di proporre una traduzione corretta, come nell'esempio n. 80.

Esempio n. 80 (Prog3_Testo8 – A/SouEr)

TO: invisible-hand-Prozessen

TA: processi di invisible-hand

PE: processi a mano invisibile

Sul piano terminologico, elemento chiave del discorso scientifico, emergono 49 errori di A/Termi/DNT (cfr. esempio n. 81), categoria impiegata per indicare termini che non necessitano traduzione. La categoria A/Termi/DNT è particolarmente rappresentata nel testo n. 5 (40 occorrenze), come si può osservare nella Tabella 9 riportata nell'allegato 4. Questi errori riguardano l'uso consolidato di termini inglesi sia nella lingua di arrivo che in quella di partenza. Per la categoria A/Termi/SemUn (cfr. esempio n. 82), relativi alla traduzione errata di termini, si registrano 44 errori. Questi dati evidenziano una presenza limitata di elementi terminologici, la cui gestione resta comunque critica.

Esempio n. 81 (Prog3_Testo8 – A/Termi/DNT)

TO: [...] hat Manuel Castells (1996) [...] den Begriff der Network Society geprägt.

TA: [...] Manuel Castells (1996) ha coniato il termine rete società [...].

PE: [...] Manuel Castells (1996) ha coniato il termine Network Society [...].

Esempio n. 82 (Prog3_Testo8 – A/Termi/SemUn)

TO: Aus der Perspektive der Akteur-Netzwerk-Theorie [...]

TA: Dal punto di vista della teoria degli attori-rete [...]

PE: Dal punto di vista della teoria dell'attore-rete [...]

Altre categorie marginali per la resa del linguaggio scientifico e che contano un numero limitato di errori sono: A/MisOrd/MW per gli errori di attribuzione semantica di elementi (34

occorrenze) e A/ForLa per gli errori di uso improprio della lingua inglese (25 occorrenze). Per quest'ultima categoria viene presentato di seguito un esempio.

Esempio n. 83 (Prog3_Testo4 – A/ForLa)

TO: Das bestätigt noch einmal eine der Grundannahmen, mit denen viele Konzepte des Neudenkens operieren [...]

TA: Questo conferma ancora una volta uno degli assunti di base con cui operano molti concetti del rethinking [...]

PE: Questo conferma ancora una volta uno degli assunti di base con cui operano molti concetti della riconsiderazione [...]

Un'ulteriore categoria che riguarda un numero contenuto di occorrenze (20 errori) è F/Sty/Idiom che comprende gli errori relativi alla resa di espressioni idiomatiche, una sfida rilevante per i sistemi di TA. Una resa sbagliata può compromettere il significato del testo di arrivo e la sua coerenza stilistica. L'esempio n. 84 ne riporta un caso.

Esempio n. 84 (Prog3_Testo1 – F/Sty/Idiom)

TO: Denn das Fehlen von kulturkreisspezifischem oder kulturübergreifendem Wissen kann zu einem Stochern im Nebel führen [...]

TA: Questo perché la mancanza di conoscenze culturalmente specifiche o interculturali può portare a una pugnalata nel buio [...]

PE: Questo perché la mancanza di conoscenze culturalmente specifiche o interculturali può portare a brancolare nel buio [...]

Nell'esempio, l'espressione idiomatica tedesca "Stochern im Nebel" è stata resa con "una pugnalata nel buio", espressione non equivalente né idiomatica nella lingua d'arrivo. La versione post-editata, "brancolare nel buio", restituisce invece l'equivalente corretto e comunemente accettato nella lingua d'arrivo. Questo tipo di errore evidenzia la difficoltà della TA nel riconoscere e interpretare unità fraseologiche fisse, nonché la tendenza, già citata, a tradurre in modo letterale, anziché proporre una traduzione basata sull'equivalenza idiomatica e sull'uso convenzionale nella lingua di arrivo.

Un numero limitato di errori si riscontra per la categoria A/MisExpl che, con 15 occorrenze, segnala le problematiche di inadeguatezza dei contenuti per il destinatario del testo di arrivo.

Esempio n. 85 (Prog3_Testo7 – A/MisExpl)

TO: Schüler/innen mit Migrationshintergrund werden [...] häufiger an Förderschulen zugewiesen [...]

TA: Le alunne e gli alunni con un background migratorio sono più spesso assegnati a scuole speciali [...]

PE: Le alunne e gli alunni con un background migratorio sono più spesso destinati alle Förderschulen, ossia scuole per studenti con bisogni educativi speciali che necessitano di un supporto didattico mirato, [...]

La proposta della TA per il termine “Förderschulen” risulta imprecisa e inadeguata per il pubblico italiano. Per garantire la comprensione del concetto nella cultura di arrivo, nella versione post-editata è stata inserita una spiegazione aggiuntiva.

Le categorie che seguono presentano un’incidenza ancora più bassa: A/MWE/Ell con 12 occorrenze; A/Mech/CharFor con 13 occorrenze; G/WF/VF con 11 occorrenze; A/Oth con 10 occorrenze; A/Mech/Capit con 5 occorrenze. Ciò suggerisce che aspetti problematici legati alla traduzione di composti con ellissi di elementi ripetuti (A/MWE/Ell), al trasferimento della formattazione (A/Mech/CharFor) o delle lettere maiuscole (A/Mech/Capit), all’uso di tempi o forme verbali (G/WF/VF) hanno un impatto minimo in questo progetto di traduzione. In riferimento agli errori di A/Oth, classificati come tali in quanto non riconducibili ad altre categorie di errori, riguardano prevalentemente la parola “tedesco” che accompagna la parola “Abstract” (cfr. esempio n. 86).

Esempio n. 86 (Prog3_Testo2_3_5_6_7_8_9 – A/Oth)

TO: Abstract (Deutsch)

TA: Abstract (tedesco)

PE: Abstract (italiano)

L’errore illustrato nell’esempio n. 86 è riconducibile a una traduzione fedele che, tuttavia, risulta inadeguata al contesto del testo di arrivo, dove l’abstract è in italiano.

Esempio n. 87 (Prog3_Testo4 – A/MWE/Ell)

TO: politische, wirtschaftliche, transport- und medientechnologische Vernetzungen

TA: interconnessioni politiche, economiche, di trasporto e di tecnologia dei media,

PE: interconnessioni politiche, economiche e tecnologiche, sia dei trasporti che dei media

Nell’esempio n. 87, viene illustrato un caso per la categoria A/MWE/Ell, in cui il sistema di TA non è riuscito a gestire correttamente la struttura ellittica del composto tedesco. In questo caso, il cosiddetto *Ergänzungsstrich* segnala l’omissione di un elemento morfologico ripetuto (“-

technologische”), che si sottintende ma si applica a più lessemi. La TA, quindi, non ha riconosciuto la funzione del trattino, ignorando conseguentemente la relazione semantica tra i due elementi.

In sintesi, gli errori direttamente riferibili alla traduzione del linguaggio scientifico ammontano a 1.165, meno di un quinto del totale (circa il 17%). Tali categorie (ad esempio A/Termi/SemUn e A/Termi/DNT per la terminologia o F/Lex/LexCho e F/Lex/Fueg per la scelta lessicale) riguardano gli aspetti caratterizzanti del linguaggio scientifico e richiedono un intervento umano mirato per assicurare un testo coerente con le convenzioni del linguaggio specialistico. Altre categorie di errore, pur non essendo specifiche delle caratteristiche del linguaggio scientifico, hanno un impatto significativo sulla qualità del testo, sulla sua accuratezza, ma anche sulla resa complessiva del linguaggio del contributo scientifico: gli errori di *Accuracy* compromettono la corretta trasmissione dei contenuti, quelli di *Fluency* incidono sulla scorrevolezza della lingua di arrivo, e quelli di *Grammar* sulla correttezza formale. Il contributo del *post-editor* risulta quindi indispensabile: l’insieme di questi interventi correttivi ha permesso di ottenere un testo finale scientificamente coerente, accurato e leggibile, in linea con i requisiti del genere testuale trattato.

6.4.2 Analisi dell’*effort* temporale e tecnico

Il totale delle sessioni per il progetto *Interculture journal* è 129:27:12, con una durata media per testo di 14:23:01. Le sessioni di una durata superiore a 01:00:00 circa sono state intervallate con pause ogni 01:30:00/01:50:00. Queste pause non sono state incluse nel conteggio della durata delle sessioni. Le indicazioni temporali per ogni singolo testo sono riportate nell’allegato 4.

Secondo il report di Post-Edit Compare, il progetto ha coinvolto nove testi, per un totale di 6.252 segmenti e 126.258 parole. La quasi totalità del materiale, pari a 6.220 segmenti, è stata generata dal sistema di TA, mentre 32 segmenti sono stati inseriti manualmente in assenza di una proposta automatica. Non si riscontrano contributi della memoria di traduzione e questo dato deve essere interpretato alla luce del funzionamento di Post-Edit Compare. Come già osservato per i due precedenti progetti di traduzione, anche in questo caso la memoria di traduzione, creata appositamente, era inizialmente vuota e si è popolata progressivamente durante lo svolgimento del *post-editing*; tuttavia, l’aggiornamento della stessa non incide sull’analisi. Post-Edit Compare confronta, infatti, unicamente la versione iniziale del progetto, salvata prima dell’avvio del *post-editing* e composta interamente dall’output della TA, con la versione finale post-editata, senza ricalcolare eventuali *matches* introdotti successivamente. Ne consegue che i testi presi in esame dal report corrispondono integralmente alla proposta della TA.

La Tabella 9 illustra le percentuali complessive rilevate nelle diverse fasce di PEMp, riportando per ciascuna di esse il numero associato di segmenti e di parole, al fine di offrire una visione dettagliata della distribuzione dei dati.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	783	10.917	17,29%
95-99%	405	10.446	16,55%
85-94%	634	17.430	27,61%
75-84%	243	6.056	9,59%
50-74%	219	5.142	8,15%
New (0-49%)	842	13.138	20,81%

Tabella 9: Distribuzione delle fasce di PEMp per progetto n. 3

L'analisi delle fasce di PEMp evidenzia complessivamente una qualità eterogenea. La fascia 100% raccoglie il 17,29% delle parole (783 segmenti, 10.917 parole), indicando che meno di un quinto del totale è stato accettato e riusato integralmente senza subire modifiche. I segmenti interessati da correzioni minime (fascia 95-99%) rappresentano il 16,55% del totale delle parole (405 segmenti, 10.446 parole), mostrando che, in una parte limitata ma non trascurabile dei casi, la TA ha prodotto un risultato di buona qualità, richiedendo solo micro-interventi. La figura seguente (35) riporta un caso di segmento sottoposto a correzione della fascia 95-99%. Nella colonna dedicata al confronto si possono osservare le correzioni effettuate rispetto alla traduzione generata automaticamente dalla macchina. Nell'ultima colonna a destra viene indicato il valore di PEMp pari al 98,08% con un'*edit-distance* di 2. Come si può osservare, si tratta di interventi estremamente mirati.

Testo di partenza	
Diese versucht, komplexe Realitäten zu beschreiben und zu verstehen, ist aber nicht mit dem identisch, was sie beschreiben will.	
Traduzione automatica	Post-editing
Questo tenta di descrivere e comprendere realtà complesse, ma non è identico a ciò che vuole descrivere.	Questa tenta di descrivere e comprendere realtà complesse, ma non è identica a ciò che vuole descrivere.
Confronto	PEMp
QuestoQuesta tenta di descrivere e comprendere realtà complesse, ma non è ~~identico~~ identica a ciò che vuole descrivere.	98.08% Edit-Dist.: 2 Max chars: 104

Figura 35: Esempio fascia di PEMp 95-99% dal testo n. 2 del progetto n. 3

La fascia 85-94% registra la percentuale più ampia, con 27,61% delle parole (634 segmenti, 17.430 parole), evidenziando che più di un quarto dell'intero progetto ha richiesto interventi di lieve

entità (cfr. Fig. 36). Ciò risulta in linea con quanto emerge nei report riguardanti i singoli testi presentati nell'allegato 4 in cui questa fascia di PEMp è sempre la più consistente, fatta eccezione per i testi n. 2 e 4.

Testo di partenza	
Auf diese Trainingsanfrage kann man mit drei Konzepten reagieren.	

Traduzione automatica	Post-editing
Potete rispondere a questa richiesta di formazione con tre concetti.	È possibile rispondere a questa richiesta di formazione con tre concetti.

Confronto	PEMp
PoteteÈ possibile rispondere a questa richiesta di formazione con tre concetti.	87.67% Edit-Dist.: 9 Max chars: 73

Figura 36: Esempio fascia di PEMp 85-94% dal testo n. 1 del progetto n. 3

Le fasce inferiori presentano risultati più limitati: la 75-84% registra il 9,59% delle parole (243 segmenti, 6.056 parole), mentre la 50-74% incide per l'8,15% (219 segmenti, 5.142 parole). Queste fasce rappresentano i segmenti coinvolti in modifiche di entità media e medio-alta. La fascia *New* (0-49%) costituisce infine il 20,81% delle parole (842 segmenti, 13,138 parole), segnalando che circa un quinto del progetto ha richiesto interventi di correzione totale o quasi totale, in seguito a un supporto non affidabile o estremamente poco affidabile della TA. La Figura 37 illustra un caso in cui il sistema di TA non ha presentato una traduzione. Questa è stata pertanto inserita manualmente in fase di *post-editing*, causando conseguentemente dei valori di PEMp e di *edit-distance* pari allo 0.

Testo di partenza	TA	Post-editing	PEMp
<cf>Im</cf> <cf>Fokus</cf> <cf>dieses</cf> <cf>Artikels</cf> <cf>steht</cf> <cf>die</cf> <cf>Frage</cf> <cf>der</cf> <cf>Wahrnehmung</cf> <cf>von</cf> <cf>kultureller</cf> <cf>bzw</cf> <cf>als</cf> <cf>solcher</cf> <cf>gedeuteter</cf> <cf>Diversität</cf> <cf>durch</cf> <cf>Lehrkräfte</cf>		<cf>Questo articolo si concentra sulla questione della percezione della diversità culturale, o interpretata come tale, da parte degli insegnanti.</cf>	0% Edit-Dist.: 0 Max chars: 143

Figura 37: Esempio della categoria di PEMp *New* dal testo n. 7 per progetto n. 3

Nel complesso, l'output automatico fornisce una base di partenza utile, ma notevolmente disomogenea. Alcune sezioni sono riutilizzabili integralmente, costituendo però una parte visibilmente ridotta del progetto nel suo complesso. Più dell'80% dei segmenti richiede modifiche. Consistenti sono gli interventi di lieve entità (85-94%), che comprendono oltre un quarto del totale dei segmenti. Non sono esclusi, tuttavia, altri tipi di interventi, quali correzioni di minima, media e medio-alta intensità, fino a comprendere rielaborazioni complete. La presenza simultanea di

segmenti accettati senza modifiche (17,29%) e di una quota consistente di *New* (20,81%) evidenzia quanto il lavoro del *post-editor* resti centrale per garantire coerenza, leggibilità e qualità complessiva in un output che è ancora lontano dal disporre della qualità desiderata.

Dal punto di vista critico, la distribuzione dei segmenti mostra che l'uso della TA, pur vantaggioso in termini di produttività per porzioni limitate, non riduce significativamente l'*effort* complessivo. Gli interventi richiesti riducono i benefici attesi dall'automazione. L'analisi conferma, quindi, l'importanza di un approccio di *post-editing* strutturato e attento per garantire un testo pubblicabile.

6.4 Riflessione conclusiva

In merito ai risultati ottenuti, in tutti e tre i progetti di ricerca la categoria *Fluency* registra il numero maggiore di occorrenze, principalmente a causa dell'elevata presenza di errori di F/Coh/InPu, che influenzano la qualità del testo scientifico. Per numero di occorrenze seguono le macrocategorie *Accuracy* e *Grammar*, quest'ultima con il numero più contenuto di errori in tutti i progetti.

Nel progetto di traduzione *Informationen Deutsch als Fremdsprache* si registrano complessivamente 6.522 errori, di cui circa il 21,5% (precisamente 1.403) sono direttamente riconducibili agli aspetti caratterizzanti del linguaggio scientifico, come le categorie relative alla traduzione della terminologia o ai fenomeni linguistici come le nominalizzazioni e l'uso delle forme impersonali. Per il progetto di traduzione *trans-kom* si rilevano complessivamente 6.893 errori e di questi circa il 18,3% (precisamente 1.262) sono errori legati direttamente alla traduzione degli elementi del linguaggio scientifico. Infine, per il progetto di traduzione *Interculture journal* si constata in totale 6.933, di cui circa il 17% (precisamente 1.165) afferiscono alle peculiarità del linguaggio scientifico. Si osserva dunque che il progetto *Informationen Deutsch als Fremdsprache* registra la quota più alta per questo tipo di errori (poco più di un quinto del totale), seguito dal progetto *trans-kom* con quasi un quinto. Il numero più basso è riscontrato nel progetto di traduzione *Interculture journal*, dove questi errori sono poco più di sesto del totale. In tutti i casi, tali errori costituiscono una componente fondamentale della traduzione degli articoli scientifici, che conferma come l'output della TA possa costituire una base di partenza utile, ma non sufficiente senza l'intervento umano a raggiungere standard qualitativi adeguati alla pubblicazione. Infatti, le analisi degli errori hanno evidenziato difficoltà dei sistemi di TA nella gestione terminologica, in quanto non impiegano sempre l'equivalente corretto nella lingua di arrivo o non riconoscono l'uso consolidato di termini in lingua inglese. Le difficoltà hanno riguardato anche la strutturazione coerente dei riferimenti bibliografici, la realizzazione delle nominalizzazioni, la corretta scelta lessicale a seconda del dominio trattato o l'impiego delle forme impersonali o di combinazioni

consolidate nell'ambito del linguaggio scientifico (*Fügungen*). Inoltre, la traduzione automatica tende a tradurre ogni segmento testuale, senza distinguere tra contenuto comunicativo e contenuto metalinguistico. In tali casi, il *post-editor* deve riconoscere che la frase non appartiene al discorso dell'autore, ma costituisce un esempio linguistico da mantenere nella lingua di partenza, al fine di preservare la coerenza metodologica e la validità scientifica del testo (cfr. ad es. esempio n. 14 in 6.2.1). Analogamente, il sistema di TA mostra limiti nella gestione delle citazioni in tedesco per le quali esistono versioni ufficiali in italiano, producendo traduzioni errate o non appropriate. È importante sottolineare che anche gli errori relativi alle altre categorie, anche se per definizione non afferiscono direttamente al linguaggio scientifico, esercitano un'influenza significativa sulla qualità del testo e sulla resa del linguaggio scientifico coinvolgendone gli elementi costituenti.

Relativamente all'*effort* temporale, il totale delle sessioni per il progetto di traduzione *Informationen Deutsch als Fremdsprache* è pari a 120:29:08; per il progetto *trans-kom* è 115:45:51; per il progetto *Interculture journal* è 129:27:12. Quest'ultimo progetto risulta il più dispendioso in termini di tempo, in linea con il numero complessivo di errori annotati (6.933). Diversamente, tra *Informationen Deutsch als Fremdsprache* e *trans-kom* la tendenza si inverte: pur presentando *trans-kom* un numero di errori maggiore (6.893 contro 6.522), il tempo totale impiegato è inferiore, segno che il *post-editing* dei testi di *Informationen Deutsch als Fremdsprache* ha richiesto un impegno maggiore. In riferimento alla lunghezza dei testi, *trans-kom* è il progetto più ampio, con 80.711 *tokens*, seguito da *Interculture journal* (76.232 *tokens*) e *Informationen Deutsch als Fremdsprache* (64.224 *tokens*). Tuttavia, la dimensione testuale non appare direttamente proporzionale al tempo impiegato per il *post-editing*. Il caso del progetto *Informationen Deutsch als Fremdsprache* è particolarmente indicativo: nonostante la minore estensione, il progetto ha richiesto un tempo di *post-editing* superiore a *trans-kom*, che presenta una lunghezza maggiore. Nel complesso, questi dati suggeriscono che la durata del *post-editing* è determinata in misura significativa dalla qualità dell'output della TA e dalla tipologia degli errori piuttosto che dalla sola quantità di testo o dal numero complessivo di errori.

Quanto all'*effort* tecnico, con il 20,65% delle parole totali, il progetto di traduzione *trans-kom* registra la percentuale più elevata di parole della TA nella fascia 100%, cioè accettate senza modifiche. Questo dato potrebbe giustificare parzialmente l'indicazione temporale più bassa rispetto agli altri due progetti. Seguono il progetto di traduzione *Interculture journal* con il 17,29% e il progetto di traduzione *Informationen Deutsch als Fremdsprache* con il 14,68%. Per la categoria *New*, cioè i segmenti sottoposti a modifiche in una percentuale da 49% al 0%, si registrano valori molto simili: il progetto *Interculture journal* registra il 20,81%; il progetto *trans-kom* il 19,11%; il progetto *Deutsch als Fremdsprache* il 18,91%. Pertanto, in tutti i casi, la quota di parole coinvolte da

interventi più consistenti è di circa un quarto del totale. Un'ulteriore analogia riguarda la fascia che registra più parole sul totale: in tutti e tre i progetti di traduzione si tratta della fascia 85-94% che registra il 25,22% per *trans-kom*, il 27,61% per *Interculture journal* e il 28,92% per *Informationen Deutsch als Fremdsprache*. Per la fascia 95-99%, i progetti *Interculture journal* e *trans-kom* registrano valori simili, rispettivamente il 16,55% e il 16,12%. Il progetto *Informationen Deutsch als Fremdsprache* registra invece un valore più basso: il 12,38%. Per la fascia 75-84% si rilevano i seguenti dati: 10,94% per *trans-kom*; 9,59% per *Interculture journal* e 15,29% per *Informationen Deutsch als Fremdsprache*. Ancora una volta i valori dei progetti *trans-kom* e *Interculture journal* mostrano risultati simili. Infine, la fascia 50-74% registra il 7,96% per *trans-kom*, l'8,15 % per *Interculture journal* e il 9,82% per *Informationen Deutsch als Fremdsprache*, mettendo in luce una somiglianza tra i valori dei progetti di traduzione in questa fascia. Si osservano dunque analogie nei tre progetti, ma anche differenze non trascurabili nei pattern di intervento e modifica. Dal punto di vista operativo, il lavoro di *post-editing* si è rivelato dunque eterogeneo e sbilanciato: accanto a segmenti pronti all'uso, sono state necessarie correzioni di varia entità, fino a rielaborazioni estese che hanno comportato un *effort* considerevole. È doveroso, tuttavia, aggiungere che i risultati delle fasce di PEMP nelle analisi complessive sono valori medi, e pertanto non consentono di cogliere le tendenze specifiche dei singoli testi, finendo talvolta per uniformare risultati di particolare rilievo. Per un quadro più dettagliato, si rimanda alle analisi dei report di Post-Edit Compare per i singoli contributi, presentate negli allegati 2, 3 e 4.

Inoltre, alla luce delle analisi esposte, è necessario precisare che, pur offrendo una base solida per riflettere sul *post-editing effort*, questi dati non ne restituiscono una visione esaustiva, in quanto non coprono tutte le dimensioni dell'*effort*. In particolare, la dimensione del *cognitive effort*, cruciale per comprendere il carico cognitivo del *post-editor*, non è stata inclusa in questa indagine a causa di limiti temporali e strumentali⁶⁵. Per quanto riguarda, invece, il *technical effort*, occorre evidenziare che i report elaborati da Post-Edit Compare considerano solo le modifiche effettuate confrontando due versioni, quella di partenza con l'output grezzo e quella finale con il testo post-editato. Essi, pertanto, non tengono in considerazione tutte le correzioni effettuate in versioni "intermedie" utili a raggiungere la versione finale post-editata, omettendo in questo modo una parte potenzialmente considerevole dell'*effort*. Alla luce di queste riflessioni, la valutazione del *post-editing effort* è da considerarsi parziale.

A quanto esposto finora, si aggiunge la componente della soggettività: l'individuazione degli errori, la formulazione di proposte alternative e il tempo impiegato sono estremamente soggettivi e possono cambiare da *post-editor* a *post-editor*. Questi aspetti possono dipendere

⁶⁵ Il *cognitive effort* è stato trattato ampiamente nel paragrafo 3.6.

dall'esperienza personale, dalla familiarità con il dominio o col genere testuale trattato, ma anche dalla propensione del *post-editor* a intervenire più o meno intensamente sui segmenti tradotti, fino allo stato d'animo e all'umore.

Ulteriori ricerche potrebbero ampliare l'indagine includendo il *cognitive effort* e considerando un corpus più ampio e diversificato di testi scientifici in ambito linguistico. Inoltre, un'analisi comparativa delle strategie di *post-editing* di diversi *post-editor* potrebbe contribuire a delineare approcci più efficaci e sostenibili nella collaborazione uomo-macchina.

Nel complesso, i risultati confermano che, nonostante i progressi della tecnologia, l'affidamento esclusivo alla TA non garantisce uniformità né adeguatezza nei testi scientifici. L'attività di *post-editing* si configura quindi come indispensabile per raggiungere una qualità pubblicabile e garantire coerenza e leggibilità complessiva, e richiede competenze linguistiche, specialistiche e metodologiche avanzate. Parallelamente, l'integrazione di Trados Studio con plugin quali Qualitivity e Post-Edit Compare si rivela utile non solo come strumento di supporto operativo, ma anche come risorsa di analisi quantitativa, in grado di fornire dati empirici oggettivi sull'interazione tra output della TA e l'intervento umano di *post-editing*. Anche in assenza di *matches*, le funzioni della memoria di traduzione, in particolare la ricerca di concordanza, così come la costituzione di un *termbase* in cui sono stati raccolti i termini man mano che venivano analizzati, hanno supportato il processo di traduzione e la coerenza terminologica, dimostrando il ruolo cruciale dei CAT tool nel processo traduttivo e nella professione del *post-editor*.

In aggiunta a quanto esposto, si segnala l'impiego di strumenti esterni utilizzati durante la fase di *post-editing* per garantire una resa accurata e coerente con il linguaggio scientifico. In particolare, Sketch Engine è stato utilizzato per analizzare il corpus in lingua italiana creato per questa fase empirica, permettendo di estrarre collocazioni e strutture lessicali tipiche della scrittura scientifica e supportando le scelte traduttive. Ulteriori strumenti hanno fornito supporto complementare: dizionari online bilingui (ad es. Sansoni Tedesco) per verificare corrispondenze lessicali tra le lingue; dizionari monolingui (ad es. Duden Online, Treccani) per chiarire sfumature semantiche e garantire un registro appropriato; banche dati di testi paralleli (Linguee, Reverso Context) per osservare l'uso contestualizzato di termini ed espressioni; motore di ricerca per la scrittura scientifica (Google Scholar) utile per verificare la frequenza d'uso e le combinazioni tipiche nel linguaggio scientifico. L'uso combinato di queste risorse ha consentito di operare in fase di *post-editing* in modo sistematico e informato, correggendo gli errori della TA, migliorando la precisione terminologica e garantendo la conformità agli standard stilistici del linguaggio scientifico.

Un'ulteriore riflessione significativa emersa riguarda l'individuazione delle tipologie di errore più ricorrenti illustrate nell'analisi presentata in questo capitolo. Tali evidenze possono costituire un punto di partenza per il miglioramento dei sistemi di TA, rappresentando un margine di manovra concreto per adattarli meglio alla traduzione del linguaggio scientifico. In questa prospettiva, il *fine-tuning* dei modelli si configura come una strategia promettente per ridurre le criticità riscontrate e rendere gli strumenti più adatti a domini specialistici. Il capitolo successivo approfondirà questa direzione, presentando un tentativo di *fine-tuning* di un sistema di TA per la traduzione del linguaggio scientifico dal tedesco all'italiano.

Studio sperimentale:

traduzione con un sistema di TA addestrato e *post-editing*

Il presente capitolo illustra lo studio sperimentale riguardante il processo di *fine-tuning* di un sistema di TA e la valutazione della resa in lingua italiana di articoli scientifici di linguistica in tedesco. Per *fine-tuning* si intende la fase di adattamento mirato di un modello di TA pre-addestrato: quest'ultimo viene inizialmente addestrato su grandi quantità di testi generici provenienti da domini eterogenei, dai quali acquisisce conoscenze linguistiche e capacità traduttive di carattere generale; il *fine-tuning* interviene successivamente per specializzarlo su un dominio specifico.

Nel presente studio tale processo si è articolato in due fasi distinte, ciascuna delle quali viene descritta di seguito, a partire dagli strumenti impiegati. Successivamente vengono illustrate le fasi metodologiche relative a ciascuna prova, dalla costituzione del *training corpus* fino alla fase di addestramento vera e propria. Per ciascuna delle due prove è stato composto un corpus che fungesse da *training set*, la cui procedura di creazione è descritta nel dettaglio. In seguito, viene presentato il procedimento di *fine-tuning* del modello. Il capitolo illustra infine le modalità di valutazione, basate sia su metriche automatiche sia su una valutazione qualitativa, quest'ultima condotta mediante un esperimento di traduzione che prevede l'individuazione, la classificazione e il *post-editing* degli errori in articoli scientifici tedeschi tradotti in italiano.

Tale struttura mira non solo a documentare con rigore il percorso sperimentale, ma anche a fornire una visione chiara, coerente e replicabile del metodo seguito, consentendo una comprensione trasparente delle scelte operative e dei risultati ottenuti.

Le attività di *fine-tuning* sono state svolte in collaborazione con il gruppo di ricerca “Mensch-Maschine Kommunikation” della ZHAW Zürcher Hochschule für Angewandte Wissenschaften, composto dalla Prof.ssa Alice Delorme Benites e dall'assistente universitario Mattia Turra.

7.1 Strumenti

Per la realizzazione dello studio empirico sono stati impiegati tre strumenti. Il primo è Transkribus, impiegato esclusivamente nella prima prova di *fine-tuning*. Il secondo strumento è LF Aligner, che trova applicazione come strumento per l'allineamento dei testi del corpus per il *training* in entrambi i tentativi di *fine-tuning*. L'ultimo strumento è OPUS-CAT MT Engine impiegato come tool per il *fine-tuning* esclusivamente nella seconda prova. Di seguito vengono descritte nel dettaglio le caratteristiche e le funzionalità di questi strumenti.

7.1.1 Transkribus

Transkribus è una piattaforma avanzata per il riconoscimento automatico dei caratteri e la gestione di testi digitalizzati. Questo tool utilizza modelli OCR (*Optical Character Recognition*) addestrati su diversi tipi di documenti e lingue, che consentono di riconoscere automaticamente il testo presente nelle immagini scannerizzate.

Transkribus è ampiamente utilizzato per decifrare i manoscritti storici, ma trova applicazione anche nei testi stampati moderni. Tra le funzionalità offerte, si individua la possibilità di creare modelli di riconoscimento del testo personalizzati, ossia modelli di intelligenza artificiale appositamente addestrati per riconoscere stili di scrittura specifici presenti nei documenti. La piattaforma consente inoltre di riconoscere tabelle e liste in vari formati, esportabili come fogli di calcolo o integrabili in database, facilitando la gestione di strutture di dati complesse.

Il tool è accessibile online al link <https://www.transkribus.org/it> e richiede la creazione di un profilo personale. L'utente può caricare i propri file e organizzarli in raccolte, beneficiando in questo modo di una gestione più agevole del lavoro. Il sistema suddivide automaticamente le pagine in righe e paragrafi, generando una rappresentazione strutturata del contenuto testuale. Una volta completato il riconoscimento automatico, la piattaforma mette a disposizione dell'utente un editor integrato, in cui è possibile correggere manualmente eventuali errori di riconoscimento e imperfezioni dovuti alla qualità delle scansioni o a layout complessi. Un aspetto particolarmente utile durante la fase di revisione manuale è la visualizzazione *side-by-side*, che mostra contemporaneamente l'immagine del documento e l'editor di testo (cfr. Fig. 38). Il sistema centra automaticamente l'immagine sulla riga in cui si sta lavorando, permettendo di individuare rapidamente la posizione nel documento e di correggere con precisione gli eventuali errori di riconoscimento.

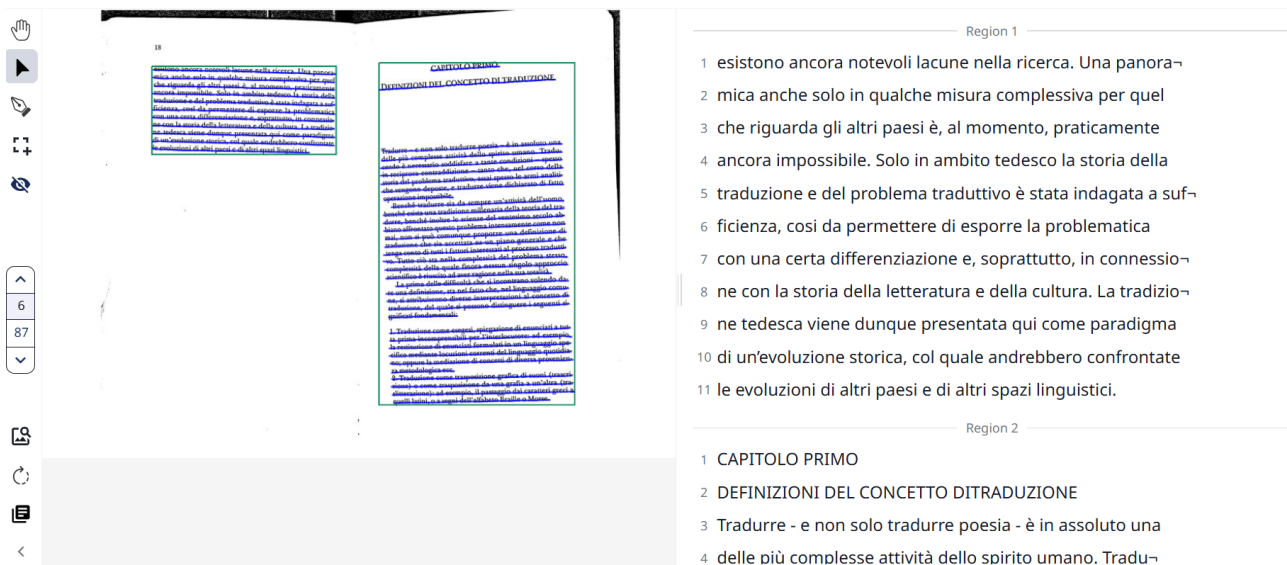


Figura 38: Visualizzazione *side-by-side* di Transkribus

Infine, Transkribus permette di esportare il testo in vari formati digitali, tra cui file .txt. Come già accennato, questo strumento è stato impiegato nel primo tentativo di *fine-tuning*, in particolare per digitalizzare i testi italiani a partire dalle scansioni di documenti cartacei. Grazie alle sue funzionalità avanzate di riconoscimento automatico e revisione manuale, Transkribus ha notevolmente facilitato la creazione del corpus parallelo tedesco-italiano.

7.1.2 LF Aligner

LF Aligner è un tool open source e gratuito, sviluppato per l'allineamento automatico di testi paralleli, disponibile per Windows, macOS e Linux al link del sito ufficiale: <https://sourceforge.net/projects/aligner/>. Si basa sull'algoritmo *hunalign* (Varga et al. 2007), un metodo ibrido che combina informazioni lessicali e informazioni basate sulla lunghezza delle frasi. In una prima fase, l'algoritmo genera una “traduzione grezza” del testo di partenza utilizzando un dizionario bilingue: questo passaggio ha lo scopo di rendere confrontabili il testo di partenza e il testo di arrivo, permettendo di individuare parole condivise tra la traduzione grezza e il “vero” testo di arrivo. Tale correlazione viene poi integrata con un secondo indice basato sul rapporto tra le lunghezze delle frasi, impiegato per valutare la loro probabile corrispondenza. La combinazione di queste due informazioni consente all'algoritmo di selezionare l'allineamento più plausibile (Varga et al. 2007).

Originariamente concepito per supportare i traduttori nella creazione di memorie di traduzione a partire da testi e dalle rispettive traduzioni, LF Aligner si è affermato come uno strumento utile anche nell'ambito della ricerca linguistica. Il software supporta diversi formati di file input (.txt, .doc, .docx, .rtf, .pdf, .html) ed è in grado di generare file output nei seguenti

formati: .tmx, .tabbed txt e .xls. È necessario specificare che i file .txt devono essere codificati in UTF-8, poiché la codifica ANSI può causare errori di lettura di alcuni caratteri speciali (ad esempio “à”, “è” o l’apostrofo).

Di seguito viene illustrato il suo funzionamento. Dopo l’installazione e l’avvio del software, il primo passaggio consiste nella scelta del file di input, selezionabile tra diversi formati compatibili (cfr. Fig. 39). In questo caso sono stati selezionati i file .txt.

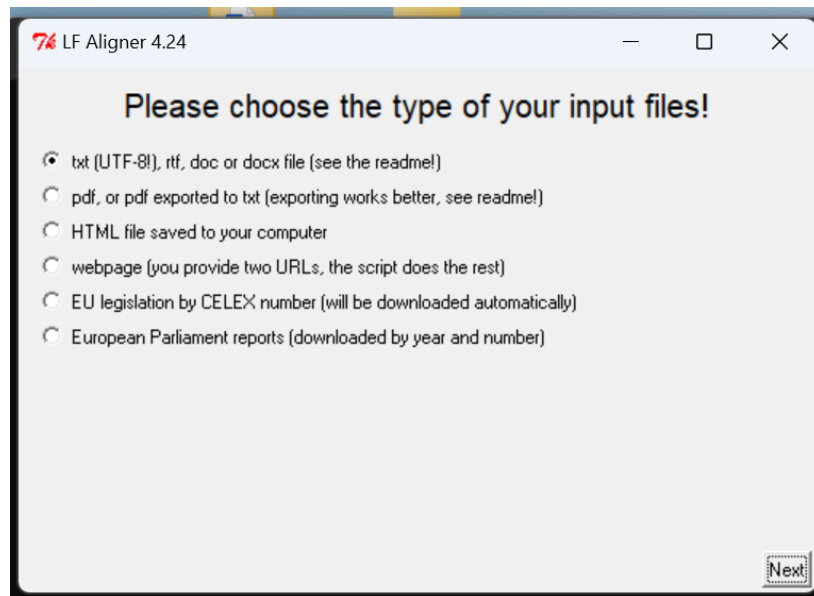


Figura 39: Finestra per la selezione del formato dei file

In seguito, l’utente deve indicare il numero di lingue coinvolte e selezionare le lingue di lavoro tramite un menu a tendina, che consente di scegliere tra oltre 180 lingue (cfr. Fig. 40).

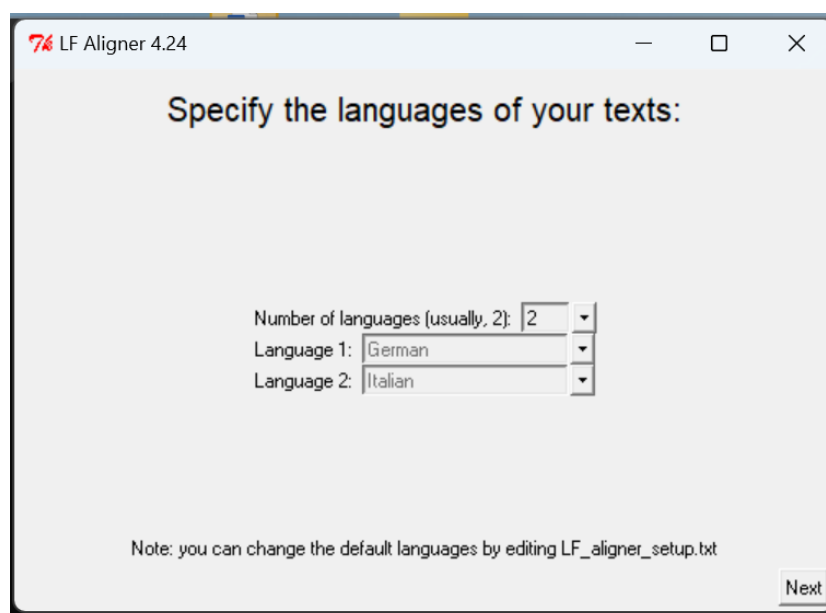


Figura 40: Finestra per l’impostazione delle lingue

Successivamente si selezionano i file da allineare: cliccando sul pulsante *Browse*, si apre una finestra dalla quale è possibile individuare e selezionare i testi su cui si desidera operare (cfr. Fig. 41).

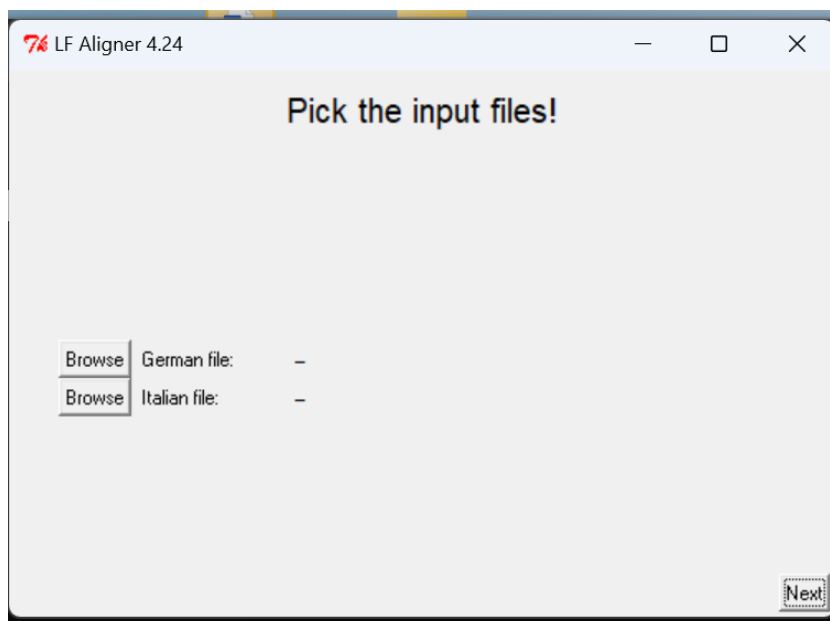


Figura 41: Finestra per la selezione dei file

LF Aligner segmenta e allinea automaticamente i testi paralleli. In caso di errori, è possibile comunque effettuare correzioni manuali tramite il *graphical editor* o generando un file .xls. Qualora non si ritenga necessario intervenire, si può saltare questa fase selezionando l'opzione *No review* (cfr. Fig. 42).

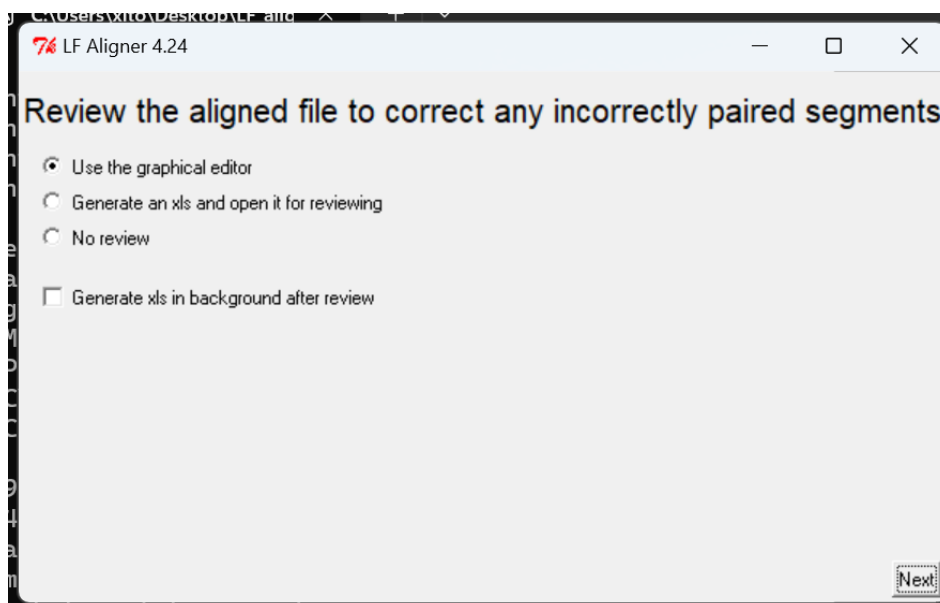
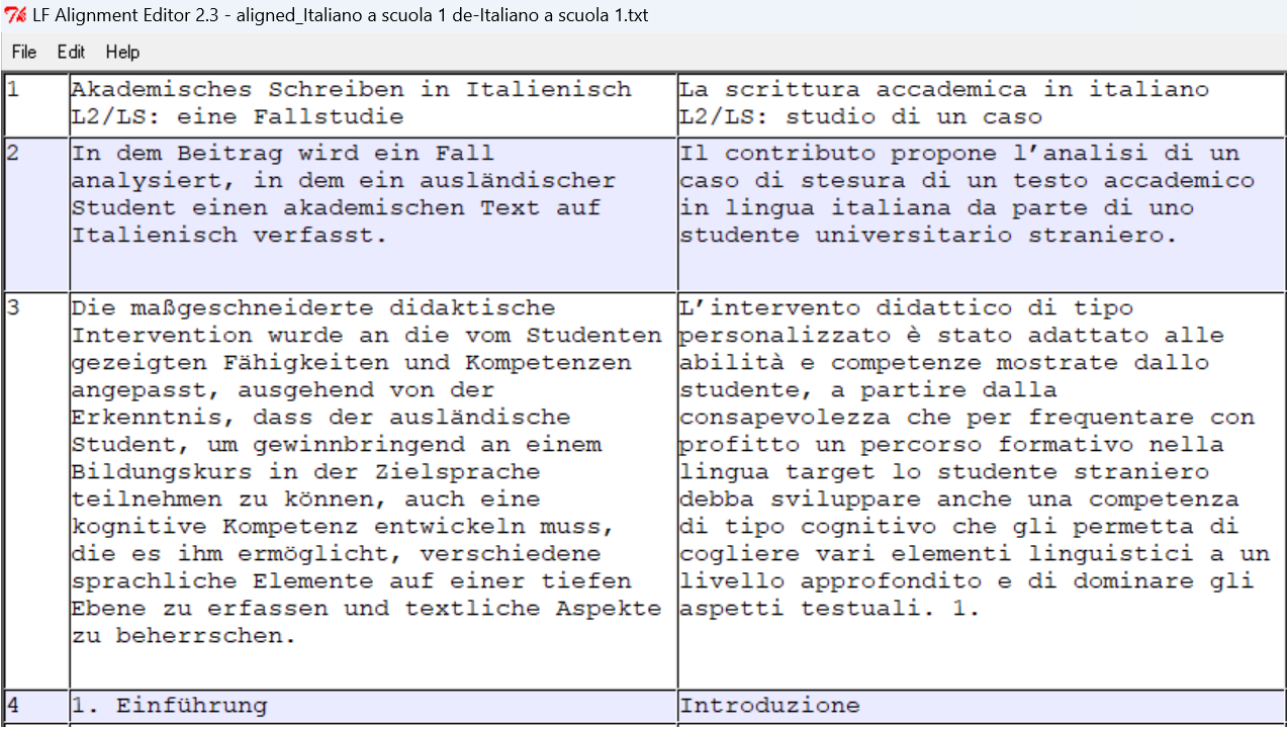


Figura 42: Finestra per opzioni di revisione

Selezionando l'opzione “use the graphical editor”, il testo viene visualizzato suddiviso in segmenti, con tre colonne principali. A partire da sinistra, la prima colonna indica il numero del segmento, la seconda mostra il testo in tedesco e la terza quello in italiano. Ogni segmento occupa una riga (cfr. Fig 43).



1	Akademisches Schreiben in Italienisch L2/LS: eine Fallstudie	La scrittura accademica in italiano L2/LS: studio di un caso
2	In dem Beitrag wird ein Fall analysiert, in dem ein ausländischer Student einen akademischen Text auf Italienisch verfasst.	Il contributo propone l'analisi di un caso di stesura di un testo accademico in lingua italiana da parte di uno studente universitario straniero.
3	Die maßgeschneiderte didaktische Intervention wurde an die vom Studenten gezeigten Fähigkeiten und Kompetenzen angepasst, ausgehend von der Erkenntnis, dass der ausländische Student, um gewinnbringend an einem Bildungskurs in der Zielsprache teilnehmen zu können, auch eine kognitive Kompetenz entwickeln muss, die es ihm ermöglicht, verschiedene sprachliche Elemente auf einer tiefen Ebene zu erfassen und textliche Aspekte zu beherrschen.	L'intervento didattico di tipo personalizzato è stato adattato alle abilità e competenze mostrate dallo studente, a partire dalla consapevolezza che per frequentare con profitto un percorso formativo nella lingua target lo studente straniero debba sviluppare anche una competenza di tipo cognitivo che gli permetta di cogliere vari elementi linguistici a un livello approfondito e di dominare gli aspetti testuali. 1.
4	1. Einführung	Introduzione

Figura 43: *Graphical editor* di LF Aligner

Questa funzionalità si è rivelata particolarmente utile nel presente studio per correggere eventuali problemi di segmentazione o allineamento, garantendo in questo modo una maggiore accuratezza del processo. Al termine della correzione, il tool consente di salvare il lavoro svolto e ottenere un file .txt contenente i segmenti allineati. LF Aligner e le sue funzionalità sono risultati significativamente appropriati in questa ricerca poiché hanno supportato la costituzione dei corpora paralleli in entrambe le prove di *fine-tuning*, consentendo la segmentazione e l'allineamento automatico dei testi in tedesco e in italiano e mantenendo al contempo la possibilità di intervenire manualmente per correggere eventuali errori.

7.1.3 OPUS-CAT MT Engine

OPUS-CAT MT Engine è un sistema di TA basato su Windows, costruito sul *framework* Marian NMT. Si tratta di un progetto *open source*, originariamente sviluppato nell'ambito del Fiskmö Project, volto a raccogliere dati paralleli finlandese/svedese da varie organizzazioni in Finlandia, e attualmente sviluppato e mantenuto come parte del progetto OPUS-MT.

OPUS-CAT MT Engine consente di utilizzare localmente una vasta gamma di modelli di traduzione neurale avanzata, offrendo traduzioni sicure e riservate direttamente sul computer dell'utente. Un ulteriore vantaggio è rappresentato dai plug-in disponibili per i principali strumenti CAT, in particolare SDL Trados Studio e memoQ, che permettono di integrare agevolmente il sistema di TA all'interno del flusso di lavoro. Inoltre, il tool può essere utilizzato anche con Wordfast come provider personalizzato, e con Memsource o XTM tramite plug-in per il browser Chrome.

Il sistema fornisce modelli di traduzione pre-addestrati per un'ampia gamma di coppie linguistiche, grazie ai progetti OPUS MT⁶⁶ e Tatoeba-Challenge⁶⁷, e permette di avviare rapidamente operazioni di traduzione o *fine-tuning* su testi già allineati. Queste caratteristiche hanno reso OPUS-CAT MT Engine particolarmente adatto al secondo tentativo di *fine-tuning* condotto nella presente ricerca. Inoltre, questo strumento è stato scelto per la sua accessibilità gratuita, la facilità d'uso e la possibilità di effettuare il *fine-tuning* senza particolari competenze tecniche, rendendolo ideale anche per utenti con esperienza limitata in *machine learning*.

Per utilizzare il software, è sufficiente scaricarlo dal sito ufficiale⁶⁸ e avviare il file OpusCatMTEngine.exe. Nella prima schermata viene presentata una lista di modelli online, tra i quali è possibile scegliere quello più adatto in base alla combinazione linguistica desiderata. Nel presente studio è stato selezionato il modello *opus-mt-de-it*, con lingua di partenza tedesco e lingua di arrivo italiano. Successivamente, premendo il pulsante "Install model locally", il modello viene installato sul computer (cfr. Fig. 44).

⁶⁶ Per ulteriori informazioni sulla lista dei modelli di OPUS MT, si veda <https://github.com/Helsinki-NLP/Opus-MT-train/tree/master/models> (ultima consultazione: 20/12/2025).

⁶⁷ Per ulteriori informazioni sulla lista dei modelli di Tatoeba-Challenge, si veda: <https://github.com/Helsinki-NLP/Tatoeba-Challenge/tree/master/models> (ultima consultazione: 20/12/2025).

⁶⁸ <https://helsinki-nlp.github.io/OPUS-CAT/> (ultima consultazione: 20/12/2025).

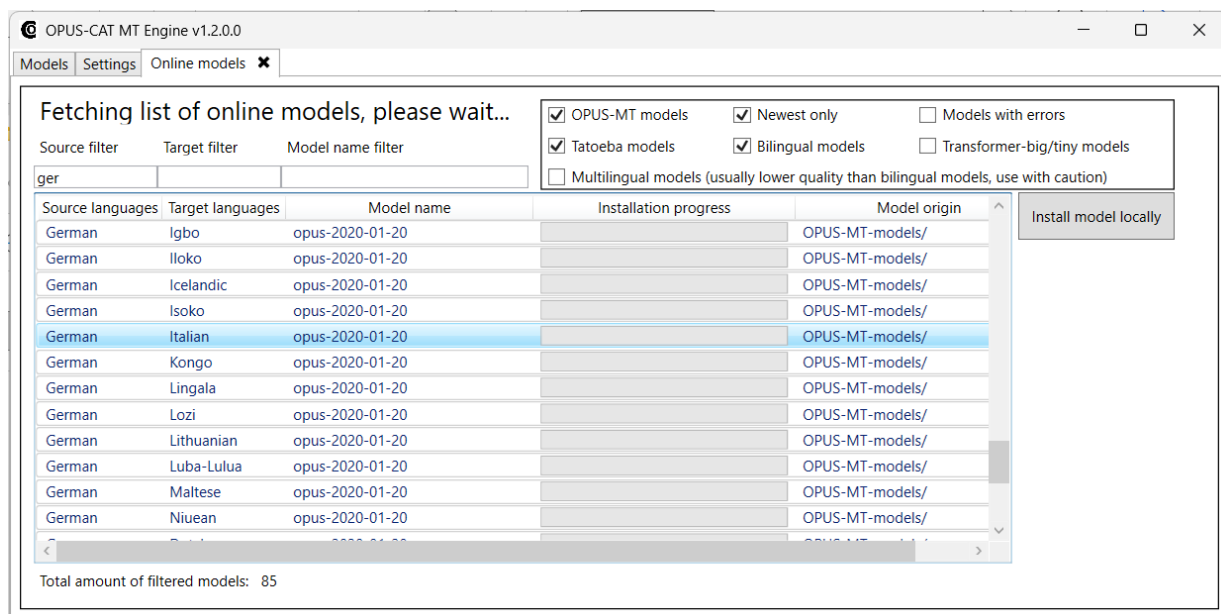


Figura 44: Schermata per la selezione di modelli in OPUS-CAT MT Engine

Una volta completata l'installazione, il modello appare nella lista dei modelli disponibili e può essere utilizzato per le traduzioni (cfr. Fig. 45).

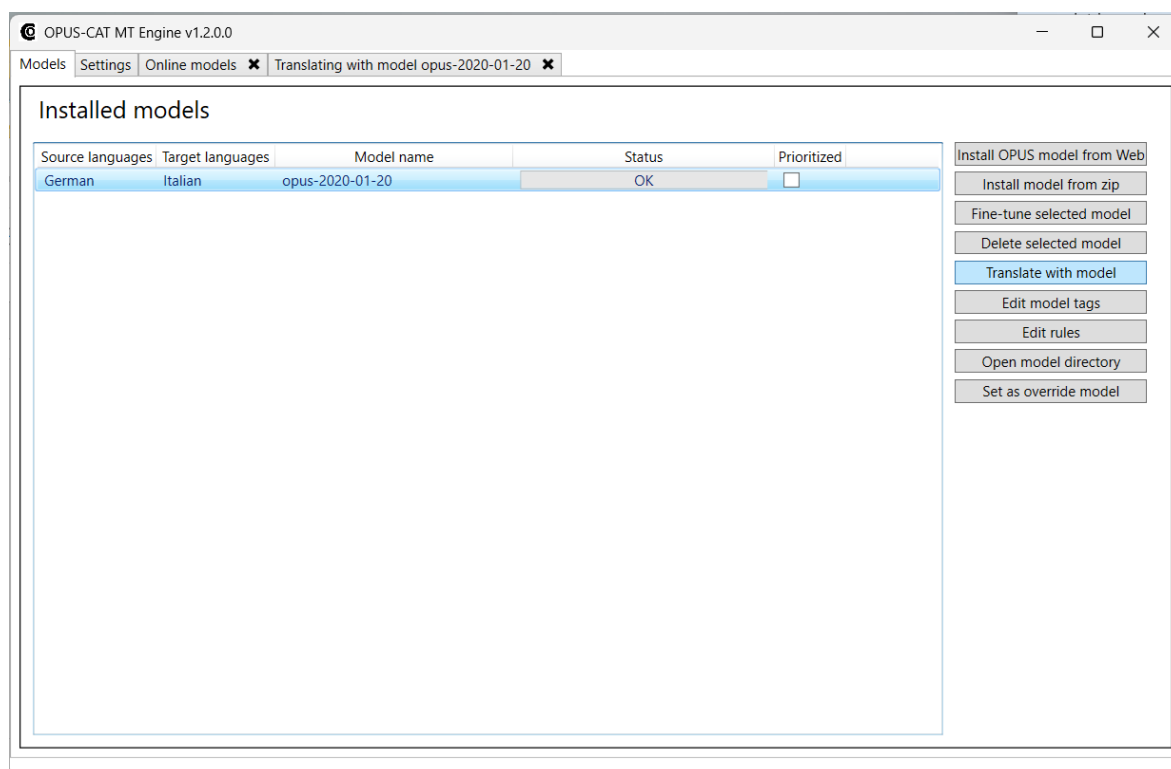


Figura 45: Schermata dei modelli installati in OPUS-CAT MT Engine

La Figura 46 mostra la schermata dedicata alla funzione di traduzione. L'interfaccia è composta da due riquadri principali: quello superiore, contrassegnato come *Source text*, consente di inserire il segmento o il testo di partenza da tradurre. Al di sotto è presente il pulsante "Translate",

che, una volta selezionato, genera automaticamente la traduzione che viene visualizzata nel riquadro inferiore, indicato come *Translation*.

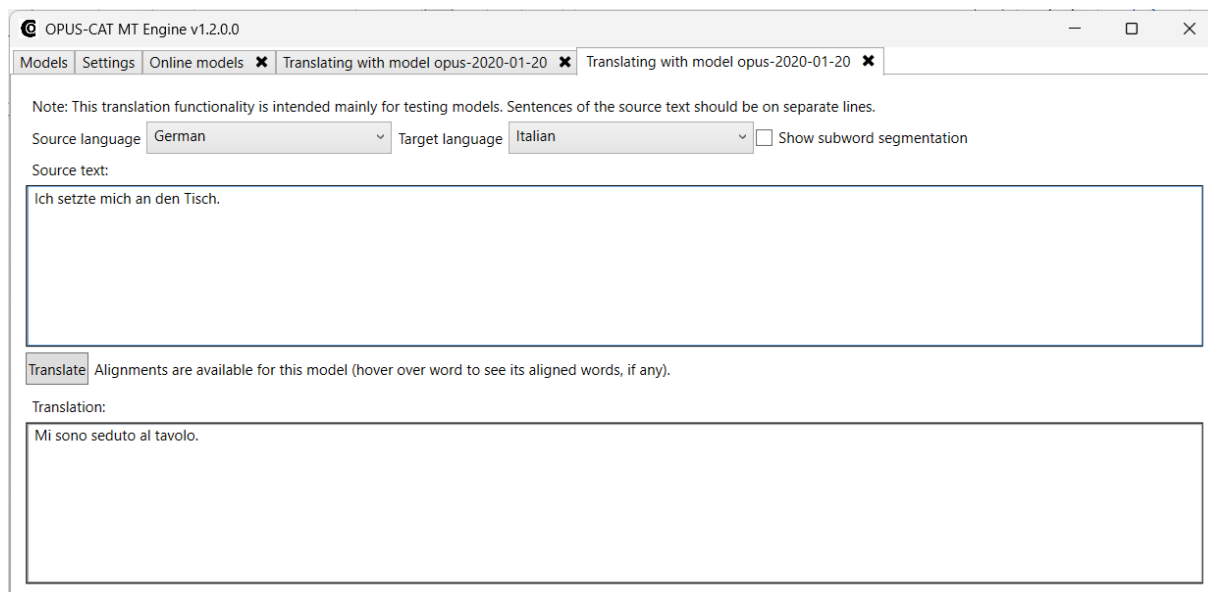


Figura 46: Schermata di traduzione in OPUS-CAT MT Engine

7.2 Prima prova di *fine-tuning*

Il primo tentativo di *fine-tuning* è stato condotto utilizzando un corpus parallelo tedesco-italiano, in cui i testi di partenza erano costituiti dai testi originali in tedesco e quelli di arrivo dalle relative traduzioni in italiano. Come già sopra accennato, con “fine-tuning” s’intende un’ulteriore fase di addestramento di un modello neurale pre-addestrato su un corpus parallelo specifico del dominio scelto (*in-domain dataset*) (Saunders 2021: 12). Si tratta di uno dei metodi di *domain adaptation* per la TA neurale (cfr. 1.3.2): il modello, inizialmente addestrato su grandi quantità di dati generici, viene esposto a un numero significativamente più ridotto di esempi appartenenti a un ambito specialistico, allo scopo di adattarne il comportamento alle caratteristiche linguistiche del dominio selezionato.

Diversi studi dimostrano che, nonostante la dimensione limitata del corpus, il *fine-tuning* consente miglioramenti significativi nella qualità della traduzione per testi dello stesso dominio, con un costo computazionale considerevolmente inferiore rispetto al riaddestramento completo del modello (Luong e Manning 2015; Servan et al. 2016). Dunque, l’obiettivo non è modificare la competenza generale del modello addestrandolo ex novo, bensì specializzarlo su un particolare ambito tematico.

Tuttavia, Saunders (2021: 12-14) segnala anche alcune criticità tipiche di questo processo: la scarsità di dati *in-domain*, il rischio di *overfitting* sul *dataset* di adattamento molto ridotto (cfr.

1.3.1.2.3) e il fenomeno di *catastrophic forgetting*, ossia il deterioramento delle prestazioni del modello su testi generici a seguito di un adattamento troppo mirato al nuovo dominio. Nel caso di questa prima prova, la difficoltà principale riguarda la limitata disponibilità di dati. Di seguito viene descritta dettagliatamente la metodologia adottata per questo primo tentativo, a partire dalla costituzione del corpus parallelo fino al processo di *fine-tuning*.

7.2.1 Costituzione del corpus

Il *fine-tuning* richiede un *dataset*, un insieme di dati di ampie dimensioni. Nel caso in esame, si è reso necessario costruire un corpus parallelo di testi scientifici in lingua tedesca e italiana.

La prima difficoltà riscontrata ha riguardato il reperimento dei testi utili a costituire il *dataset*. Inizialmente sono stati ricercati articoli scientifici che trattassero le tematiche della traduzione, della comunicazione interculturale e del tedesco come lingua straniera e che fossero disponibili in versioni parallele, l'originale in lingua tedesca e la traduzione in italiano. La ricerca non ha prodotto risultati significativi. Successivamente, l'indagine si è concentrata su volumi, scritti originariamente in tedesco, che fossero stati tradotti in italiano e che trattassero le tre aree tematiche già citate. Anche in questo caso la ricerca si è rivelata complessa, in quanto è stata riscontrata una diffusa mancanza di letteratura specialistica germanofona tradotta in lingua italiana afferente alle tematiche di interesse. A conclusione di questa indagine, sono stati individuate tre opere di riferimento in lingua tedesca:

1. Beaugrande, Robert Alain de, e Wolfgang Ulrich Dressler. 1981. *Einführung in die Textlinguistik*. Max Niemeyer Verlag.
2. Bühler, Karl. 1965. *Sprachtheorie: die Darstellungsfunktion der Sprache*. 2., Unveränd. Aufl. Fischer.
3. Apel, Friedmar. 1983. *Literarische Übersetzung*. Sammlung Metzler 206. J.B. Metzler.

con le relative traduzioni in lingua italiana:

1. Beaugrande, Robert Alain de, e Wolfgang Ulrich Dressler. 1984. *Introduzione alla linguistica testuale*. Tradotto da S. Muscas. Il mulino.
2. Bühler, Karl. 1983. *Teoria del linguaggio: la funzione rappresentativa del linguaggio*. Tradotto da S. Cattaruzza Derossi. A. Armando.
3. Apel, Friedmar. 1993. *Il manuale del traduttore letterario*. 1. Ed. A cura di Emilio Mattioli. Tradotto da G. Rovagnati. I saggi di testo a fronte 6. Guerini.

I primi due volumi hanno come tematica la linguistica generale, il terzo volume la traduttologia. Non è stato possibile reperire opere tradotte in lingua italiana pertinenti alle aree della

comunicazione interculturale e della didattica del tedesco come lingua straniera o, più in generale, della didattica delle lingue straniere.

Per poter eseguire un processo di *fine-tuning* è indispensabile disporre di dati in formato digitale. Poiché i volumi erano disponibili unicamente in formato cartaceo, si è quindi dovuto procedere con la loro digitalizzazione. I testi in lingua tedesca sono stati acquisiti tramite scanner dotato di funzione OCR, in grado di riconoscere automaticamente il contenuto e di restituirlo in formato testuale. Questa operazione è stata svolta con uno strumento che consentiva l'elaborazione della lingua tedesca, ma non di quella italiana. Per i testi in lingua italiana si è adottata quindi una procedura differente: le scansioni sono state effettuate senza usufruire della funzione OCR e successivamente elaborate tramite Transkribus. Come già illustrato, Transkribus, oltre a generare automaticamente i file di testo, offre un editor che permette di intervenire manualmente sul contenuto riconosciuto. Già in questa fase è stato dunque possibile effettuare un primo intervento di *data cleaning*, correggendo errori e imperfezioni derivanti dal processo di riconoscimento.

Dalla digitalizzazione si sono poi ottenuti file in formato .txt: nel caso dei testi italiani, generati direttamente da Transkribus e scaricabili dalla piattaforma; nel caso dei testi tedeschi, estrapolati a partire dalle digitalizzazioni in formato .pdf. Questi file hanno rappresentato la base per le fasi successive di allineamento e preparazione del corpus parallelo.

Una volta ottenute le versioni digitali, i testi in tedesco e in italiano sono stati allineati mediante LF Aligner. L'allineamento è stato infine rifinito tramite un'ulteriore fase di *data cleaning*, volta a correggere eventuali disallineamenti o residui errori di riconoscimento dei caratteri. Al termine del processo, è stato prodotto un corpus parallelo costituito da 11.091 segmenti.

7.2.2 *Fine-tuning* manuale

Nel primo tentativo di *fine-tuning*, per la selezione del modello pre-addestrato, cioè il modello di partenza, sono stati presi in considerazione requisiti specifici. In primo luogo, è stata considerata fondamentale la disponibilità della combinazione linguistica tedesco-italiano; inoltre, si è tenuto conto dell'accessibilità al corpus utilizzato per il pre-addestramento del modello e della possibilità di operare con modelli di dimensioni ridotte, adeguati a un esperimento preliminare. In questo senso, i modelli Marian, su cui si basa OPUS-MT, rappresentano una soluzione consolidata e affidabile per l'attività di adattamento, grazie alle loro dimensioni contenute e alla loro disponibilità in modalità *open source*.

Per implementare concretamente il processo di *fine-tuning* è stato necessario fare ricorso a un *framework* che consentisse di gestire in modo efficiente il caricamento, l'addestramento e la valutazione dei modelli di TA. A tale scopo è stato utilizzato PyTorch, un *framework* di *deep*

learning open source sviluppato da Facebook AI Research (FAIR) nel 2016. PyTorch è una libreria scritta in Python che fornisce strumenti per la costruzione e l'addestramento di reti neurali complesse. Nella presente ricerca PyTorch riveste un ruolo cruciale in quanto costituisce l'infrastruttura sottostante alla libreria Transformers di Hugging Face, che mette a disposizione modelli pre-addestrati disponibili per essere caricati e adattati a nuovi compiti tramite *fine-tuning*⁶⁹. Il modello selezionato, opus-mt-de-it, è un modello Marian pre-addestrato sulla combinazione linguistica tedesco-italiano⁷⁰. Prima del *fine-tuning*, il modello era già stato addestrato su diversi corpora paralleli generici disponibili in OPUS⁷¹, come avviene per i modelli appartenenti al progetto OPUS-MT (Tiedemann et al. 2024). Questo pre-addestramento consente al modello di acquisire conoscenze linguistiche di base, che vengono successivamente adattate tramite il *fine-tuning* al dominio specifico scelto, in questo caso articoli scientifici di linguistica. Il modello è stato caricato attraverso la classe MarianModel della libreria Transformers di Hugging Face. Per il *preprocessing* è stato utilizzato il corrispondente MarianTokenizer, utile a gestire la suddivisione delle frasi in *tokens*.

In seguito all'allineamento, il corpus parallelo è stato salvato in un file in formato CSV, contenente le coppie di segmenti paralleli (frase in tedesco e la corrispondente traduzione in italiano). Il file è stato caricato tramite la libreria datasets di Hugging Face, che consente una gestione ottimizzata dei dati per compiti di *machine learning*. Successivamente, il *dataset* è stato suddiviso in due sottoinsiemi: il *training set*, pari al 90% dei dati complessivi, utilizzato per l'addestramento del modello, e il *test set*, pari al 10%, riservato alla valutazione del modello su dati non visti in precedenza.

La configurazione del *training* è stata definita attraverso la classe Seq2SeqTrainingArguments della libreria Transformers, che consente di specificare i principali iperparametri del processo. Questi comprendono: tasso di apprendimento (*learning rate*) pari a $2e-5$ ⁷², dimensione del *batch* (*batch size*) di 16 ⁷³ (sia per il *training* che per la valutazione), *weight decay* di $0,01$ ⁷⁴ e un totale di 3 epoche di addestramento⁷⁵. La valutazione del modello è stata

⁶⁹ Per maggiori informazioni su Hugging Face si veda <https://huggingface.co/> (ultima consultazione: 20/12/2025)

⁷⁰ <https://huggingface.co/Helsinki-NLP/opus-mt-de-it> (ultima consultazione: 20/12/2025)

⁷¹ I corpora paralleli di OPUS sono disponibili al seguente link: <https://opus.nlpl.eu/> (ultima consultazione: 20/12/2025)

⁷² Il *learning rate* indica quanto velocemente il modello aggiorna i suoi pesi durante l'addestramento. Un valore troppo elevato potrebbe indicare che il modello omette delle soluzioni ottimali; un valore eccessivamente basso comporta un addestramento troppo lento.

⁷³ Il *batch size* segnala quanti esempi vengono processati simultaneamente. In questo caso, il modello guarda 16 coppie di frasi alla volta.

⁷⁴ Il *weight decay* è una tecnica di regolazione che aggiunge una penalizzazione ai pesi troppo grandi per evitare fenomeni di *overfitting*. $0,01$ è il valore che "limita" la crescita dei pesi.

⁷⁵ Le epoche di addestramento costituiscono il passaggio completo del modello su tutti i dati di *training*. In questo caso si segnalano 3 epoche, cioè il modello ha visto l'intero corpus tre volte durante il *fine-tuning*.

effettuata al termine di ciascuna epoca utilizzando metriche di generazione di sequenze. L'effettivo processo di *fine-tuning* è stato condotto tramite la classe Seq2SeqTrainer, che automatizza i passaggi del *training*. Infine, il modello *fine-tuned* è stato valutato rispetto al *validation set*. La valutazione è stata condotta sia attraverso la misura della *sequence-level loss*⁷⁶, che quantifica l'errore medio di predizione, sia tramite il calcolo del punteggio BLEU⁷⁷ (Papineni et al. 2002), specifico per i compiti di TA.

Questo approccio di *fine-tuning* ha illustrato una metodologia semplice, ma comprovata ed efficace per adattare un modello pre-addestrato al compito traduttivo specifico da svolgere. Sebbene questo tentativo di *fine-tuning* sia stato portato a termine con successo, sono emerse delle criticità, alcune delle quali riscontrate già prima dell'addestramento specifico della macchina. La prima, già menzionata, ha riguardato la difficoltà nel reperimento del materiale idoneo a costruire un *dataset* di dimensioni sufficienti: sebbene la quantità di dati raccolta potesse essere considerata accettabile, è presumibile che un corpus più ampio avrebbe consentito il conseguimento di risultati più significativi ed efficaci. Un'ulteriore problematica era connessa alla natura stessa dei testi selezionati: la maggior parte di essi risaliva agli anni Ottanta, con due eccezioni, una pubblicazione del 1965 e un'altra del 1993. Ne derivava che il linguaggio scientifico contenuto nei volumi sarebbe potuto risultare in parte datato e non più pienamente in linea con le convenzioni linguistiche attuali. Tale aspetto risultava particolarmente rilevante considerando che l'intento della fase sperimentale era condurre un piccolo esperimento di traduzione con il sistema *fine-tuned* su testi già analizzati nella fase empirica precedente, appartenenti a un arco temporale dal 2013 al 2022. Pertanto, un potenziale scarto nelle convenzioni linguistiche tra il corpus di addestramento e i testi presi in esame avrebbe potuto incidere negativamente sulla validità dei risultati. Inoltre, il *training* ha impiegato diversi giorni, un tempo significativo, considerando l'hardware a disposizione.

Nonostante le criticità sopra descritte fossero in larga parte già note prima dell'avvio del *fine-tuning*, si è scelto di procedere comunque con una prima fase sperimentale a carattere esplorativo, volta ad acquisire una comprensione più concreta del funzionamento del modello *fine-tuned* e a verificare la fattibilità dell'approccio. I limiti emersi durante questa fase hanno suggerito la necessità di esplorare soluzioni alternative, potenzialmente più efficienti e meglio adattabili al compito traduttivo considerato, come illustrato nella sezione seguente.

⁷⁶ La *sequence-level loss* misura quanto il modello "sbaglia" nella previsione della sequenza corretta rispetto alla traduzione di riferimento.

⁷⁷ Per maggiori informazioni sulla metrica di valutazione BLEU, si rimanda al paragrafo 2.2.2.1.

7.3 Seconda prova di *fine-tuning*

A seguito della prima prova di *fine-tuning*, caratterizzata da tempi di addestramento considerevoli a causa delle risorse hardware disponibili e da dati bilingui raccolti insufficienti e/o inadeguati per quantità e natura, si è deciso di adottare una strategia alternativa volta ad aumentare e migliorare il materiale di *training*.

Per raggiungere tale obiettivo si è fatto ricorso al concetto di corpus sintetico. Per *synthetic data* si intende un insieme di dati generati artificialmente che riproduce le caratteristiche del dato reale (Little 1993). Una strategia semplice ma empiricamente efficace per ottenere un corpus parallelo sintetico consiste nella *back-translation*: a partire da un corpus monolingue originale nella lingua di arrivo si genera automaticamente una versione nella lingua di partenza tramite traduzione automatica, ottenendo così coppie parallele di segmenti di arrivo e di partenza. Sennrich et al. (2016: 88) definiscono il processo di *back-translation* come “an automatic translation of the monolingual target text into the source language”, sottolineando come tale procedura consenta di impiegare nel corpus testi autentici nella lingua di arrivo, in quanto prodotti da parlanti umani, affiancati a testi di partenza di natura sintetica ma funzionali alla costituzione del corpus. L’efficacia di questo approccio nella formazione di un *training corpus* risiede proprio nella possibilità di preservare l’autenticità della lingua di destinazione: sebbene i dati di partenza del corpus siano sintetici, la presenza di testi target originali permette al modello di apprendere strutture e convenzioni linguistiche aderenti all’uso naturale della lingua di arrivo, rendendo teoricamente possibile l’ottenimento di un output traduttivo di qualità ottimale.

I corpora sintetici sono emersi subito come strumenti efficaci nel *Natural Language Processing* (NPL), in quanto permettono di aggirare il problema della scarsità di dati e quindi di migliorare le performance di modelli nello svolgimento di diversi compiti. In tal senso sono diversi gli studi che attestano l’uso dei corpora sintetici. Tra questi si individua lo studio di Ueffing et al. (2007), in cui si presenta il *training* di un sistema di TA statistica mediante l’uso di una piccola quantità di dati, ampliata traducendo dati monolingui in un’altra lingua che fungeva da lingua di arrivo del *training set*. Questo è stato uno dei primi studi a permettere un adattamento dei modelli di traduzione a un dominio specifico senza impiegare necessariamente dati bilingui appartenenti all’ambito di interesse, ma dati di partenza monolingui autentici, come negli studi descritti di seguito.

Un altro studio estremamente rilevante è quello condotto da Bertoldi e Federico (2009), in cui è stata esaminata la *domain adaptation* in sistemi di TA statistica come metodo funzionale per ottenere miglioramenti significativi. Impiegando un sistema *baseline*, è stato generato un corpus bilingue sintetico partendo da dati monolingui in un dominio specifico, scritti nella lingua di arrivo,

e traducendoli nella lingua di partenza del corpus. In modo analogo, Bojar e Tamchyna (2011) hanno sviluppato la tecnica di *reverse self-training* che sfrutta dati monolingui nella lingua di arrivo per ampliare il vocabolario del modello in questione. Lo studio ha dimostrato che tale tecnica migliora la capacità del sistema di produrre traduzioni grammaticalmente corrette anche in lingue morfologicamente complesse, risultando utile anche in contesti con pochi dati paralleli disponibili. Infine, vale la pena citare lo studio condotto da Sennrich et al. (2016) che dimostra evidenti miglioramenti nelle performance di sistemi di TA neurale in seguito all'addestramento con dati sintetici generati tramite *back-translation*. Nell'ambito di questo studio sono stati osservati miglioramenti della *fluency* e nelle performance della traduzione di domini specifici, mostrando anche la superiorità di questa tecnica rispetto a metodi alternativi come l'uso di frasi di partenza vuote o fittizie (*empty or dummy source sentences*), in cui il dato di partenza è nullo o contiene un *token* segnaposto, consentendo al modello di imparare a generare correttamente la lingua di arrivo anche in assenza di contesto di partenza.

Nel presente studio le intuizioni di Sennrich et al. (2016) sono state messe in pratica costruendo un corpus in cui i testi di arrivo italiani autentici sono stati accoppiati a testi di partenza tedeschi sintetici ottenuti tramite traduzione automatica inversa, in modo da preservare la naturalezza e l'autenticità della lingua di arrivo e disporre di materiale sufficiente nella lingua di partenza, utile per il *training*, anche se sintetico. Le fasi successive, che hanno portato allo svolgimento del processo di *fine-tuning*, sono illustrate in maniera dettagliata di seguito.

7.3.1 Costituzione del corpus

Le sezioni seguenti espongono nel dettaglio i diversi passaggi che hanno portato alla costituzione del corpus sintetico impiegato come *training set* di questa seconda prova di *fine-tuning*.

7.3.1.1 Scelta dei testi

Per la costituzione del corpus sono stati impiegati gli articoli scientifici in italiano già selezionati per il corpus in lingua italiana creato per il precedente studio empirico (cfr. Capitolo 5). Si tratta di contributi estratti da diverse riviste scientifiche *open access* e riconosciute dall'ANVUR come riviste scientifiche di Classe A. Le riviste sono state scelte sulla base degli ambiti tematici trattati, in modo tale da garantire la maggiore affinità possibile con quelli delle riviste impiegate per il corpus di lingua tedesca già utilizzato precedentemente, i cui temi erano la didattica delle lingue straniere, la traduzione e la comunicazione interculturale.

Le riviste selezionate per la costituzione di questo corpus sono le seguenti: *Educazione interculturale*, *EL.LE*, *inTRAlinea*, *Italiano a scuola*, *Italiano a stranieri* – *Edilingua*, *Italiano*

LinguaDue, *Lingua culture mediazioni*, *Lingue e linguaggi*, *LTO – Lingue e testi di oggi*, *Mediazioni*, *Rivista internazionale di tecnica della traduzione*. Le edizioni da cui sono stati selezionati i contributi appartengono a un arco temporale che va dal 2013 al 2023. Il numero degli articoli selezionati ammonta a 200, per un totale di 1.188.417 *tokens* (cfr. 5.1). La tabella seguente illustra le distribuzioni quantitative degli articoli tra le diverse riviste.

Nome della rivista	Numero di articoli
Educazione interculturale	18
EL.LE	18
inTRAlinea	20
Italiano a scuola	7
Italiano a stranieri – Edilingua	30
Italiano LinguaDue	27
Lingua culture mediazioni	9
Lingue e linguaggi	18
LTO – Lingue e testi di oggi	10
Mediazioni	14
Rivista internazionale di tecnica della traduzione	29

Tabella 10: Testi per il *training set*

I file contenenti i testi sono stati scaricati dalle pagine online delle riviste in formato .pdf. Successivamente sono stati trasformati in file Word, che a loro volta sono stati sottoposti a una fase di esportazione per ottenere i testi in formato .txt. Durante le fasi di conversione si sono riscontrate alcune problematiche che hanno richiesto delle operazioni di *data cleaning*, illustrate di seguito.

7.3.1.2 *Data cleaning* manuale

Inizialmente è stato eseguito un *data cleaning* manuale applicato in alcuni casi ai file in formato Word e in altri ai file in formato .txt, al fine di correggere errori derivanti dalla formattazione errata o dal trasferimento non corretto di caratteri, analoghi a quelli già individuati e descritti negli interventi di *pre-editing* dello studio empirico (cfr. 5.3.2). Le principali tipologie di intervento effettuate sono state:

- Spazi aggiuntivi tra le parole: alcuni file presentavano spazi multipli tra le parole. Questa problematica è stata corretta direttamente sui file Word utilizzando la funzione “Sostituisci”: nella barra “Trova” sono stati inseriti due o tre spazi consecutivi, mentre nella barra “Sostituisci con” è stato inserito un singolo spazio. Cliccando su “Sostituisci tutto”, tutti gli spazi in eccesso

sono stati automaticamente sostituiti con una sola spaziatura (cfr. Fig. 47). Tale criticità ha riguardato in particolar modo i testi della rivista *inTRAlinea*.

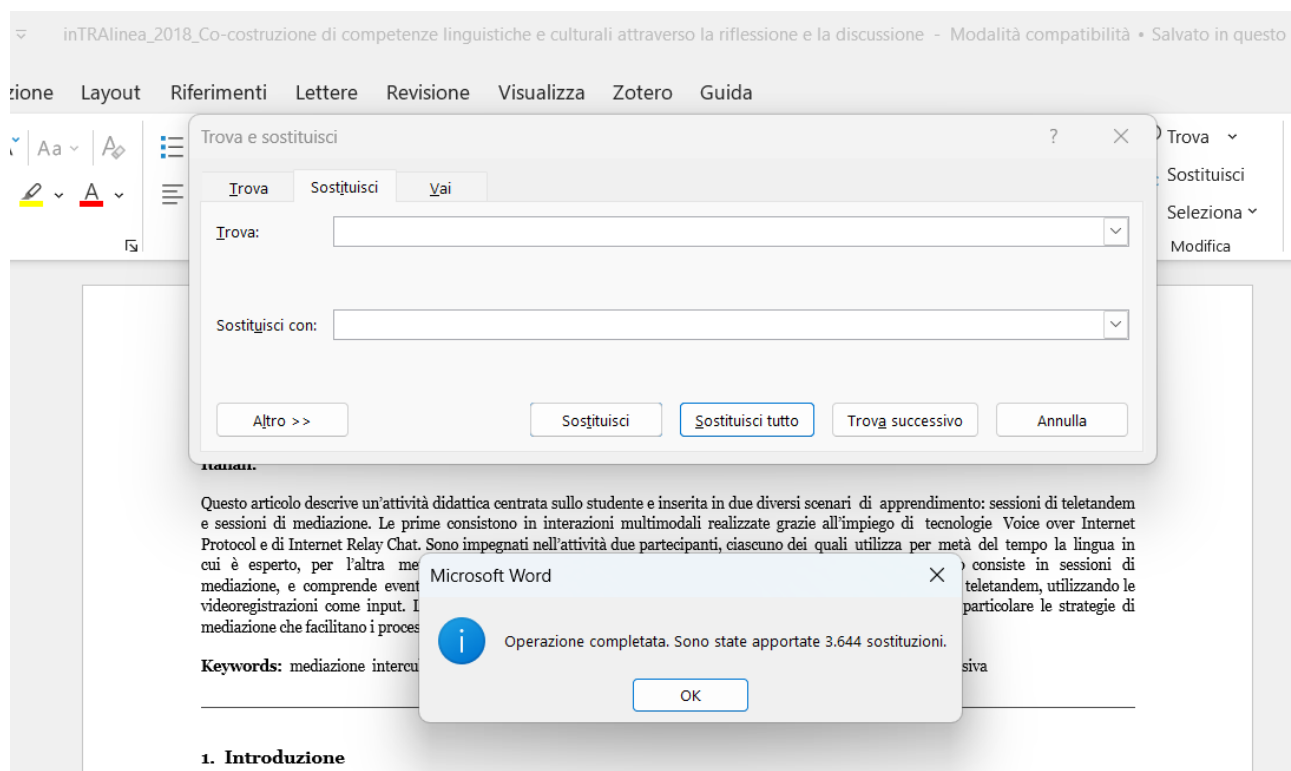


Figura 47: Problematica spazi aggiuntivi tra le parole

- Interruzioni di riga non corrispondenti alla fine della frase: alcune frasi nel file .txt risultavano “spezzate” a causa di interruzioni di pagina o di sezione nel file .pdf originale (cfr. Fig. 48). Questo ha portato alla separazione di parole appartenenti alla stessa frase e alla creazione di nuovi paragrafi indesiderati. Come mostrato nella Figura 49, la frase si sviluppa a cavallo di due pagine diverse del file .pdf, risultando quindi frammentata.

2.1.1. Il test di ingresso

Come anticipato, la funzione principale del test di ingresso è quella di rendere consapevole lo studente delle sue difficoltà nella lingua madre e di offrire al docente un quadro attendibile dei problemi e del livello degli studenti per trarre indicazioni utili su come costruire il percorso didattico. A differenza degli anni precedenti in cui il test d'ingresso era piuttosto articolato (si veda Gatta & De Santis, 2013; Gatta & La Forgia, in stampa) perché finalizzato anche ad una indagine più approfondita sulle competenze grammaticali e testuali delle matricole in uscita dalle scuole superiori, nell'anno qui considerato si è proposto un test molto semplice, visto l'uso soprattutto “interno” e strumentale del test.

Figura 48: Problematica interruzione di riga nel file .txt

2.1.1. Il test di ingresso

Come anticipato, la funzione principale del test di ingresso è quella di rendere consapevole lo studente delle sue difficoltà nella lingua madre e di offrire al docente un quadro attendibile dei problemi e del livello degli studenti per trarre indicazioni utili su come costruire il percorso didattico. A differenza degli anni precedenti in cui il test d'ingresso era piuttosto articolato (si veda Gatta & De Santis, 2013; Gatta & La Forgia, in stampa) perché finalizzato anche ad una

page 2

Printed with Pdfcrowd.com

indagine più approfondita sulle competenze grammaticali e testuali delle matricole in uscita dalle scuole superiori, nell'anno qui considerato si è proposto un test molto semplice, visto l'uso soprattutto "interno" e strumentale del test.

Figura 49: Problematica interruzione di riga nel file .pdf

- Uso del punto interrogativo al posto dell'apostrofo: questo errore è stato riscontrato esclusivamente in quattro articoli della rivista *Italiano LinguaDue* e ha visto l'uso del punto interrogativo in sostituzione dell'apostrofo per i casi di elisione (cfr. Fig. 50).

comunicative. Il(i) modo(i) in cui la complessità e la ricchezza espressiva dello spazio linguistico del nostro Paese si riverberano nell'italiano appreso dagli stranieri ha stimolato un viepiù vivo interesse della comunità accademica, a

Figura 50: Uso del punto interrogativo nell'elisione

Tale problematica ha riguardato anche le parole che richiedono l'apostrofo a fine parole a causa di un troncamento, come mostra la figura seguente.

accoglie la richiesta d'aiuto e suggerisce il corrispettivo in italiano argomento (turno 10), ripetuto sequenzialmente da Ithania a mo? di conferma (turno 11). Dal turno

Figura 51: Uso del punto interrogativo nel troncamento

Le operazioni descritte hanno consentito di eliminare gran parte delle incongruenze più evidenti, riducendo in maniera significativa la probabilità di errori residui nei file finali. L'intervento manuale è stato riservato esclusivamente agli errori la cui correzione non comportava un dispendio eccessivo di tempo o risorse cognitive. Nei casi in cui la complessità delle correzioni risultava maggiore, si è invece optato per un processo di *data cleaning* automatico, come descritto nella sezione successiva.

7.3.1.3 Data cleaning automatico

Durante la fase di *data cleaning* manuale è stato individuato un errore ricorrente in numerosi testi, imputabile alla formattazione del file .pdf di origine. Come mostrato nella Figura 52, l'anomalia si manifestava nella suddivisione impropria di parole, spezzate con l'inserimento di un trattino (-) seguito da uno spazio.

propósito di quali contenuti relativi alla tecnologia devono essere inclusi nella formazione dei traduttori. L'obiettivo dell'articolo è quello di presentare

Figura 52: Problematica suddivisione di parole

Questo fenomeno è riconducibile all'interruzione di riga presente nel documento .pdf originale, che è stato trasferito nel file convertito in formato .txt. L'errore è stato rilevato in 34 testi appartenenti alle seguenti riviste: *Italiano a stranieri – Edilingua*, *Lingue culture e mediazioni* e *Rivista internazionale di tecnica di traduzione*. Considerata la frequenza elevata e la diffusione estesa del problema, un'eventuale correzione manuale avrebbe comportato un dispendio di tempo e risorse significativo. Per questo motivo si è optato per un intervento di *data cleaning* automatico, implementato mediante uno script in linguaggio Python (cfr. Fig. 53), in grado di rilevare e correggere sistematicamente tali occorrenze.

```
File "<stdin>", line 1
  replace_pattern_in_folder(default_path, pattern,"")
IndentationError: unexpected indent
>>>
>>>
>>> def replace_pattern_in_folder(folder, pattern, replace):
...     """
...     Finds all txt-files in the folder provided, reads them, passes the text through the regex
and writes the cleaned up version to another file.
...     """
...     # fetch all txt files
...     all_txt_files = glob.glob(f"{folder}/*.txt")
...     #iterate through list of txt files
...     for file in all_txt_files:
...         #reads original file
...         with open(file, "rt", encoding="utf-8") as fin:
...             #seperates filename and extionsion (e.g. .txt)
...             base_name, ext = os.path.splitext(file)
...             # creates new file name for the cleaned up document
...             new_file_name = base_name + "_clean.txt"
...             #creates new file with provided name
...             with open(new_file_name, "w", encoding="utf-8") as fout:
...                 # iterates through every line of the original file
...                 for line in fin:
...                     # writes the cleaned up version of the text to the new file
...                     fout.write(re.sub(pattern, replace, line))
...
>>>
>>> if __name__ == "__main__":
...     main()
```

Figura 53: Codice in Python per correggere l'uso improprio di “– ”

Lo script si basa sull'impiego di espressioni regolari (*regex*), attraverso le quali viene identificato lo schema “– ” (trattino seguito da spazio) quando questo compare tra due lettere. In tal modo, le parole spezzate vengono automaticamente ricomposte nella loro forma corretta. Il programma scansiona tutti i file .txt presenti in una cartella, applica la sostituzione riga per riga e genera una nuova versione dei documenti “puliti”, riducendo drasticamente la presenza dell'errore e garantendo una maggiore uniformità testuale.

7.3.1.4 Traduzione, segmentazione e allineamento

A partire dai file .txt ottenuti, è stata effettuata la traduzione automatica dei dati dall'italiano al tedesco impiegando il sistema di TA DeepL, nella versione a pagamento DeepL Pro Advanced. La scelta di questa versione è motivata da tre fattori fondamentali: la maggiore sicurezza dei dati, l'assenza di un limite di lunghezza per i testi da tradurre e la possibilità di elaborare documenti completi. In questa fase, i file .txt in lingua italiana sono stati tradotti automaticamente in tedesco. A conclusione del processo, DeepL ha restituito i corrispondenti file .txt contenenti i testi tradotti.

Una volta ottenuti i materiali in entrambe le lingue, si è proceduto con la segmentazione e l'allineamento a livello di frase della traduzione tedesca con i testi originali in italiano, impiegando il software LF Aligner. Come input sono stati scelti i file .txt e successivamente si è impostato *German* come prima lingua, e quindi come lingua di partenza, e *Italian* come seconda lingua, e cioè come lingua di arrivo, in linea con la direzione di traduzione oggetto dello studio sperimentale. L'elaborazione è stata condotta tramite il *graphical editor* del programma, che ha permesso di verificare la correttezza della segmentazione e l'allineamento, intervenendo manualmente nei casi di errore. La Figura 54 illustra un errore di segmentazione verificatosi nel segmento 861, dove due frasi sono state unite erroneamente in un unico segmento.

361	Das Einzige, was der Übersetzer tun kann, ist abzuschätzen, inwieweit der ungewohnte oder sogar unbekannte Begriff auf die Notwendigkeit zurückzuführen ist, eine gewisse Neuheit in den Beobachtungen zu beschreiben, die noch nicht in medizinischen Handbüchern, geschweige denn in Fachwörterbüchern enthalten ist. ²² In einem solchen Fall ist die durch die wörtliche Übertragung fremder Worte eingeführte Fremdheit, die so weit wie möglich mit Respekt vor der eigenen Sprache erfolgt, aus diagnostischen und therapeutischen Gründen und nicht aus sprachlichen Gründen gerechtfertigt.	L'unica cosa che il traduttore può fare è valutare quanto il termine poco noto o addirittura sconosciuto sia dovuto alla necessità di descrivere una certa novità nelle osservazioni, ancora non inclusa nei manuali di medicina, né tanto meno nei dizionari specialistici. ²² In tale caso, la stranezza importata attraverso la letterale trasposizione delle parole altrui, nei limiti del possibile effettuata col rispetto per la propria lingua, è giustificata da motivi diagnostici e terapeutici e non linguistici.
362	3. ZEIT/FREQUENZ	3. TEMPO /FREQUENZA
363	Ein weiterer wichtiger Wert in jedem Bericht ist die Zeit, eine Dimension, in der physiologische und pathologische Phänomene auftreten.	Un altro valore importante in ogni referto è il tempo, una dimensione in cui avvengono fenomeni fisiologici e patologici.
364	Der klinische Bericht ist eine Summe von Beobachtungen, die in einem	La relazione clinica è una somma di osservazioni effettuate in un limitato

Figura 54: Problema di segmentazione in LF Aligner

L'editor si è rivelato fondamentale non solo per la correzione manuale di eventuali errori di segmentazione e allineamento, ma anche per effettuare un ultimo intervento di *data cleaning* manuale, volto a rimuovere elementi superflui residui nel testo. Al termine di questa fase, sono stati

ottenuti 28.510 segmenti allineati. Il risultato dell'allineamento, cioè il nuovo corpus di addestramento specifico, è stato quindi esportato in un file .txt contenente l'insieme dei segmenti definitivi.

7.3.2 *Fine-tuning* assistito

Disponendo di un *dataset* più ampio e consistente, si è proceduto all'implementazione del secondo tentativo di *fine-tuning*. Diversamente dal *fine-tuning* manuale, questo approccio è stato condotto utilizzando l'OPUS CAT MT Engine. Si tratta di uno strumento progettato principalmente per il *fine-tuning* dei modelli OPUS-MT mediante memorie di traduzione, ma supporta anche l'elaborazione di file .txt. Grazie agli interventi precedentemente effettuati, i dati necessitavano solo di essere adattati al formato richiesto, così da poter essere letti e processati correttamente. In particolare, è stato necessario fornire due file specifici per lingua contenenti le frasi allineate.

Sebbene sia disponibile un file di impostazioni modificabile, le possibilità di configurazione del *fine-tuning* tramite questo strumento risultano limitate, non permettendo quindi modifiche più sostanziali.

Per avviare la fase di *fine-tuning*, è necessario in primo luogo selezionare il modello che si intende adattare nella finestra *Models*. In questo caso è stato scelto il modello opus-mt-de-it, già impiegato per la prima prova di *fine-tuning*: si tratta di un modello Marian pre-addestrato per la combinazione linguistica tedesco-italiano sulla base di diversi corpora paralleli generici disponibili in OPUS. Una volta selezionato, è sufficiente cliccare sul pulsante “fine-tune selected model” (cfr. Fig 55).

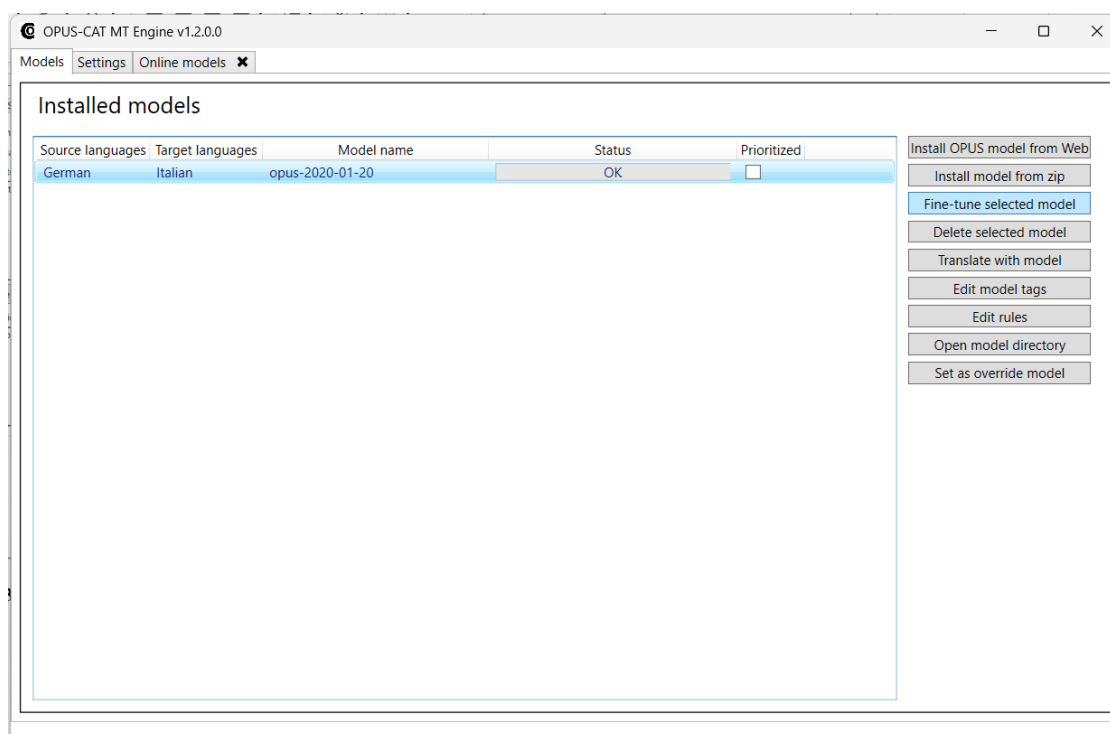


Figura 55: Schermata modelli installati in OPUS-CAT MT Engine

Si apre quindi una finestra in cui l'utente ha la possibilità di caricare i file che intende usare per il *fine-tuning* (cfr. Fig. 56). Durante il *fine-tuning* con OPUS-CAT, i *validation file* svolgono la funzione di monitoraggio della qualità del modello e dei progressi del processo. Si tratta di coppie di file (*source file* e *target file*) che non vengono utilizzate per l'addestramento, ma esclusivamente per testare periodicamente le traduzioni generate dal modello. Al termine di ogni epoca di addestramento, OPUS-CAT traduce i *validation file* di partenza e confronta le traduzioni in italiano prodotte con i dati di riferimento presenti nei *validation file* di arrivo in italiano, calcolando metriche di similarità che permettono di verificare se il modello sta migliorando nella produzione di traduzioni, avvicinando progressivamente le traduzioni prodotte a quelle di riferimento. L'"avvicinamento" consiste pertanto in un incremento della similarità misurata da tali metriche, ossia in un controllo automatico e statistico basato sul confronto fra segmenti e relativo alla misura in cui la traduzione automatica riproduce correttamente le sequenze di parole rispetto ai dati di riferimento (cfr. 7.3.3), e non in una valutazione linguistica qualitativa (cfr. 7.3.4). Come impostazione predefinita, i file di validazione vengono selezionati automaticamente e casualmente a partire dai file utilizzati per il *fine-tuning*. Tuttavia, come mostrato nella Figura 56, è possibile specificare manualmente *validation file* separati, selezionando l'opzione "Select separate files" nella sezione *Validation files*. In questo caso si è deciso di selezionare l'impostazione predefinita.

È inoltre possibile assegnare un nome al nuovo modello adattato indicandolo nel riquadro “Fine-tuned model label”. Dopo aver configurato tutte le impostazioni desiderate, il *fine-tuning* può essere avviato cliccando sul pulsante “Fine-tune”.

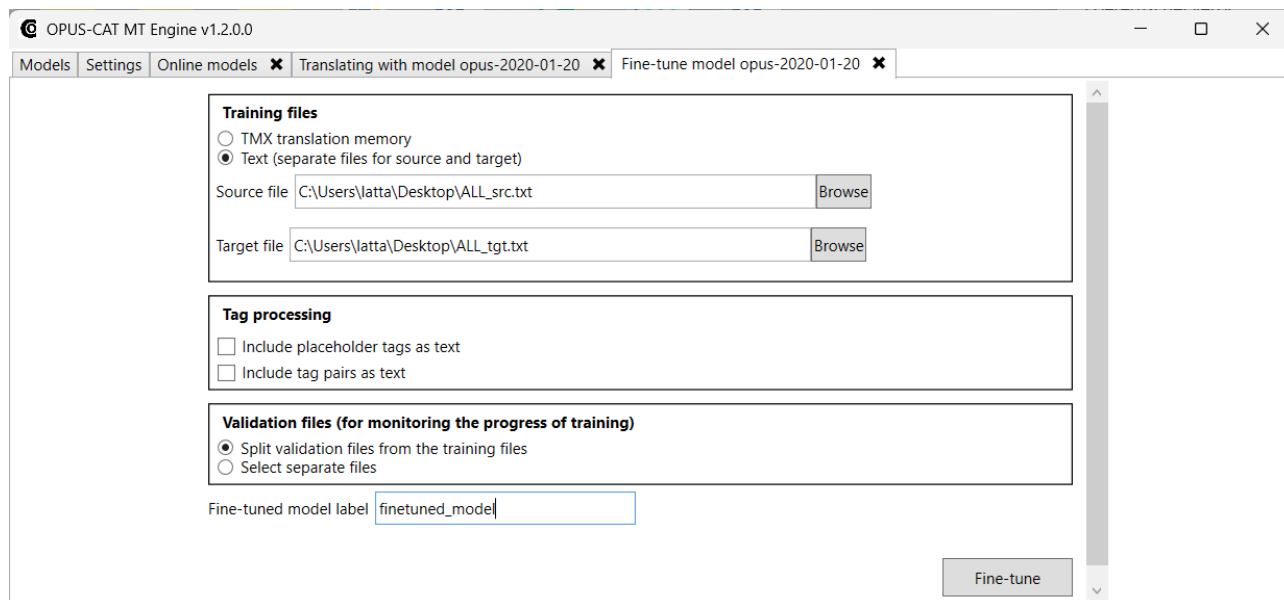


Figura 56: Schermata *fine-tuning* in OPUS-CAT MT Engine

Poiché il *fine-tuning* può richiedere tempi piuttosto lunghi, è utile monitorarne l’andamento nella scheda “Models” per assicurarsi che il processo si svolga correttamente. La colonna “Status” mostra una descrizione testuale della fase in corso (ad esempio *Preprocessing training files*), sovrapposta a una barra di avanzamento che si riempie progressivamente in verde. Una volta completata la fase di *preprocessing*, lo *Status* cambia in *Fine-tuning* e viene visualizzata una stima della durata residua del processo, calcolata sulla base della velocità registrata fino a quel momento, con un margine di imprecisione (cfr. Fig. 57).

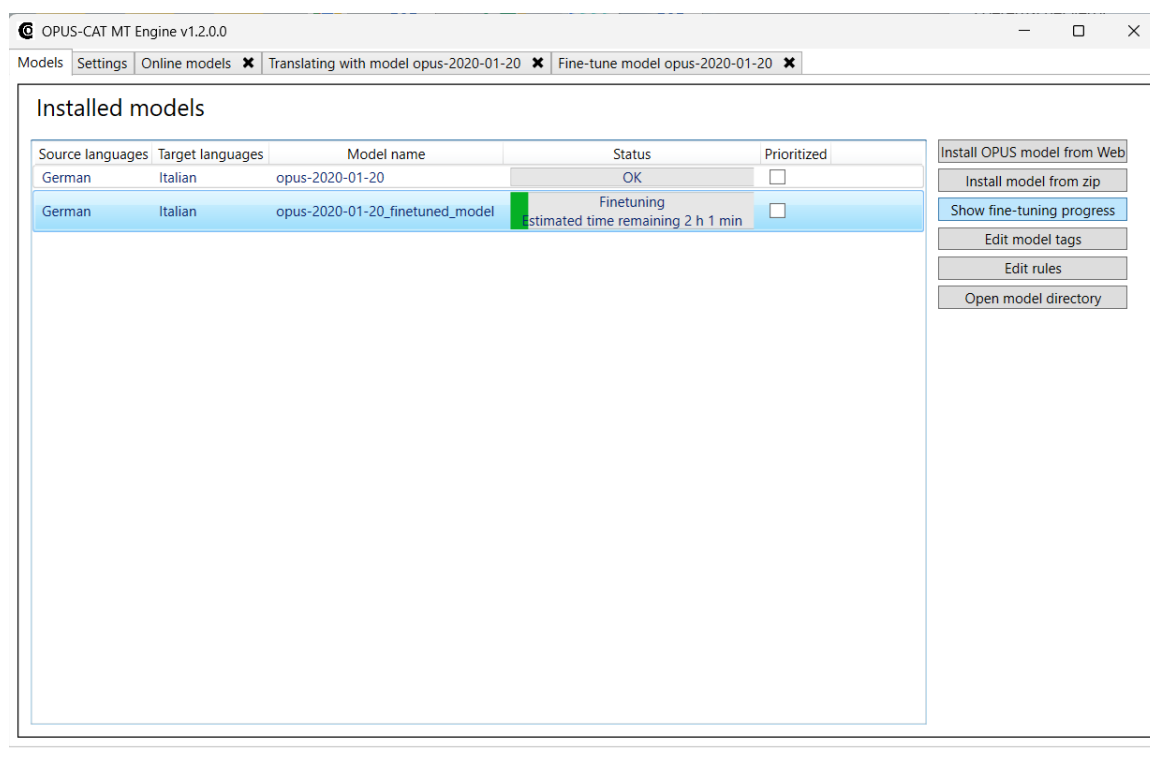


Figura 57: Schermata *fine-tuning* in OPUS-CAT MT Engine

Selezionando il modello in fase di adattamento e cliccando su “Show fine-tuning progress”, è possibile osservare come il modello performa rispetto al *validation set* man mano che il *fine-tuning* procede. Nella stessa scheda sono mostrati i punteggi sul *validation set in-domain*, costituito dai *validation file* forniti precedentemente e appartenenti allo stesso dominio del *dataset* utilizzato per il *training*, e sul *validation set out-of-domain*, estratto dal corpus Tatoeba e composto principalmente da traduzioni provenienti da domini eterogenei. Pertanto, il *set out-of-domain* serve esclusivamente a monitorare la capacità del modello di generalizzare su testi non specifici del dominio di addestramento, mentre il *fine-tuning* viene effettuato solo sui dati *in-domain*. All’inizio del processo, i punteggi *out-of-domain* possono risultare simili o leggermente superiori a quelli *in-domain*, poiché il modello pre-addestrato possiede già conoscenze linguistiche generali. Con il procedere del *fine-tuning*, è atteso un miglioramento progressivo dei punteggi *in-domain*, che riflette l’adattamento al dominio specifico, mentre i punteggi *out-of-domain* tendono a rimanere stabili o a mostrare lievi diminuzioni.

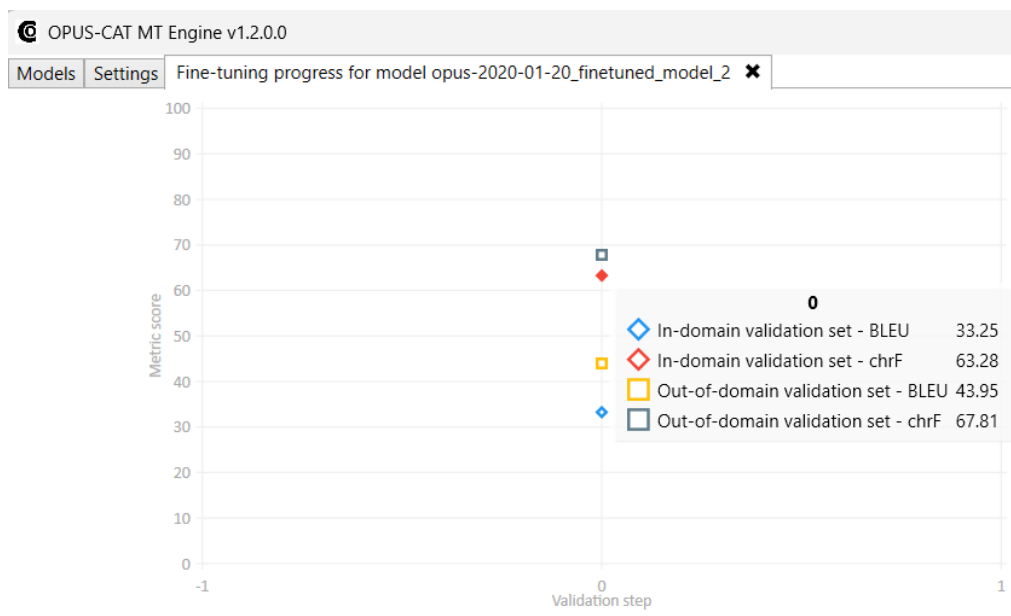


Figura 58: Schermata *Fine-tuning progress* in OPUS-CAT MT Engine

Nella Figura 58 sono mostrati i risultati per le metriche BLEU e chrF. Come già menzionato (cfr. 2.2.2.1), BLEU (*Bilingual Evaluation Understudy*), misura la sovrapposizione di n-grammi di parole tra traduzioni generate e frasi di riferimento (Papineni et al. 2002). chrF, invece, valuta la similarità a livello degli n-grammi di caratteri (Popović 2015). I punteggi sono mostrati nell'asse Y, mentre l'asse X rappresenta gli step di validazione (*validation step*). In questo caso lo step è 0, in quanto mostra i valori relativi a un *fine-tuning* appena iniziato.

I risultati indicano che le prestazioni del modello risultano leggermente migliori sui dati *out-of-domain*, suggerendo che il dominio di validazione interno potrebbe essere più complesso e/o che il *fine-tuning* iniziale non ha ancora ottimizzato completamente i parametri per il dominio specifico. Tuttavia, entrambe le metriche mostrano valori coerenti con una buona qualità di traduzione automatica per un modello neurale di medie dimensioni, con miglioramenti attesi sui dati *in-domain* man mano che il *fine-tuning* procede.

7.3.3 Valutazione automatica

Al fine di ampliare la base di analisi, sono state adottate ulteriori metriche, includendo nuovamente il punteggio BLEU e simulando un *fine-tuning* manuale con il medesimo sistema e *training set*. Anche in questo caso l'addestramento del modello è stato condotto per tre epoche, con valutazioni intermedie effettuate ogni circa 0,8 epoche. La metrica automatica BLEU, già introdotta nel paragrafo precedente, confronta le traduzioni generate dal modello con le frasi di riferimento considerate corrette, misurando la sovrapposizione di n-grammi (sequenze di parole) tra la traduzione automatica e quella di riferimento. In questo contesto, quindi, la qualità della traduzione

non riguarda lo stile o la scorrevolezza linguistica, ma indica quanto la traduzione prodotta riproduce correttamente le sequenze di parole attese, migliorando la corrispondenza lessicale rispetto alla frase di riferimento dei *validation file*, cioè il segmento in italiano.

La Figura 59, riguardante l'evoluzione del BLEU score, mostra un miglioramento costante nel corso delle epoche: 21,43 alla prima valutazione (epoca 0,8), 22,34 alla seconda (epoca 1,6), 22,83 alla terza (epoca 2,4) e 23,09 alla quarta (epoca 3). L'incremento progressivo del punteggio indica che il modello ha appreso gradualmente a generare traduzioni più accurate rispetto alle frasi di riferimento in italiano, migliorando la corrispondenza lessicale e la coerenza degli n-grammi. Il trend positivo, privo di flessioni o decrementi, suggerisce una fase di apprendimento stabile, senza *overfitting* o segni di saturazione, quindi senza mostrare rallentamenti o blocchi nell'apprendimento.

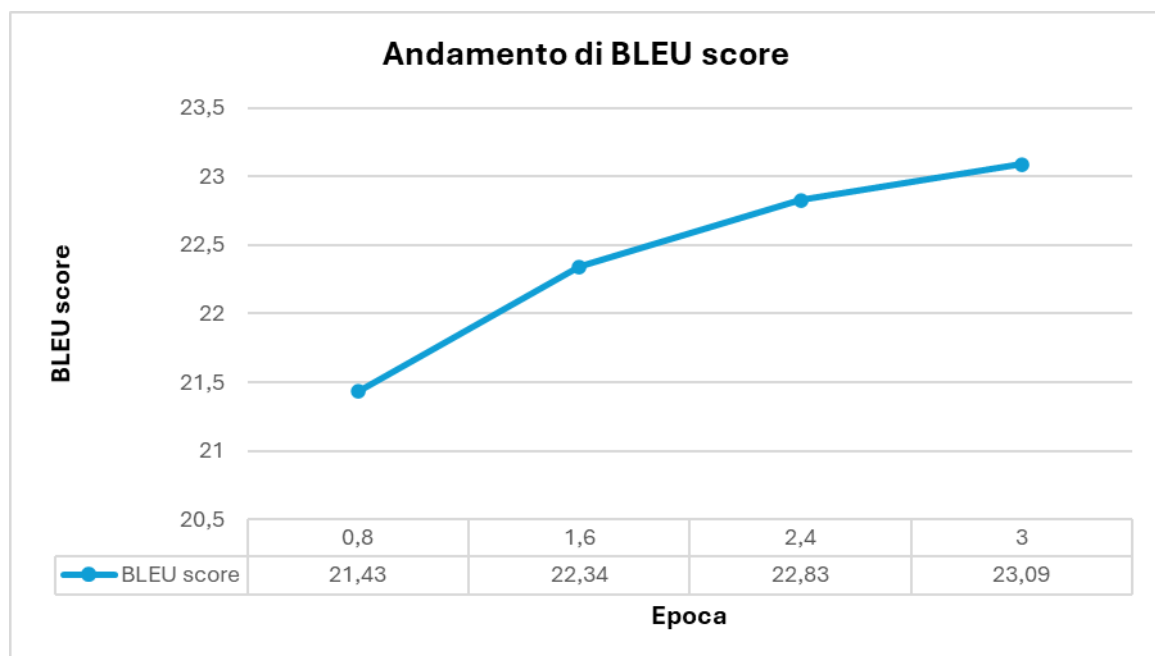


Figura 59: Andamento di BLEU score

Oltre a BLEU, sono state considerate la *training loss* e la *validation loss*. La *training loss* rappresenta l'errore del modello sui dati utilizzati per l'addestramento (*training set*), misurato come la distanza tra le predizioni del modello e le sequenze di riferimento. La *validation loss*, invece, indica la prestazione del modello su un insieme di dati separato (*validation set*), che non viene utilizzato durante l'addestramento, permettendo di valutare la capacità del modello di generalizzare su dati non visti e di rilevare eventuali fenomeni di *overfitting*. Nel caso dei modelli Marian, sia la *training loss* che la *validation loss* sono calcolate mediante la *cross-entropy*, una funzione di perdita standard nei modelli di TA neurale (Saunders 2021). La *cross-entropy* misura quanto le probabilità assegnate ai *tokens* predetti dal modello *fine-tuned* differiscono dai *tokens* corretti presenti nelle

frasi di riferimento: quindi un valore di *training loss* o *validation loss* più basso indica che il modello sta imparando a prevedere correttamente i *tokens* della sequenza di riferimento.

Come illustrato nella Figura 60, entrambe le curve mostrano una generale tendenza alla diminuzione nel corso dell'addestramento, segnale che il modello riduce progressivamente l'errore sia sul *training set* sia su quello di validazione.

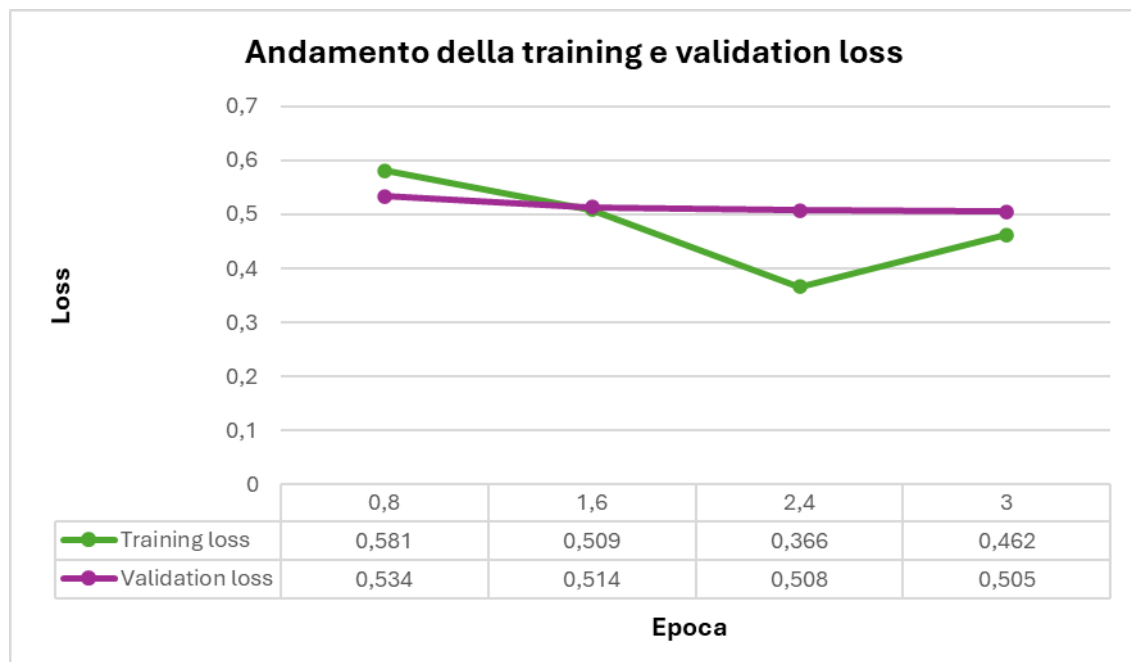


Figura 60: Andamento della *training* e *validation loss*

La distanza contenuta tra i due valori per l'intera durata del processo suggerisce una buona capacità di generalizzazione e l'assenza di *overfitting*. La lieve oscillazione osservabile nelle ultime epoche rientra nella normale variabilità statistica e non compromette la lettura complessiva dell'andamento.

Nel complesso, la riduzione costante della *loss* e il parallelo incremento del punteggio BLEU confermano la regolarità e l'efficacia del processo di apprendimento, indicando che il modello sta affinando la sua capacità di associare in modo corretto i segmenti in lingua di partenza (tedesco) con le corrispondenti sequenze in lingua di arrivo (italiano). In altre parole, il modello migliora progressivamente nella generazione di traduzioni che risultano più vicine alle frasi di riferimento, con una maggiore sovrapposizione di n-grammi rispetto alle traduzioni attese. I miglioramenti osservati restano tuttavia contenuti, in linea con le dimensioni ridotte del *training set* e con le caratteristiche del sistema di TA utilizzato.

7.3.4 Traduzione, annotazione degli errori e *post-editing*

I miglioramenti evidenziati nelle analisi di valutazione automatica fanno riferimento a dati meramente quantitativi. Per un'analisi orientata qualitativamente, la presente ricerca sperimentale prevede una fase di valutazione del nuovo addestramento mediante l'individuazione, l'annotazione e la correzione degli errori nelle traduzioni generate dal modello *fine-tuned*, al fine di esaminare in modo più approfondito le prestazioni del sistema di TA nella traduzione del linguaggio scientifico tedesco in italiano in ambito linguistico.

La valutazione umana è stata condotta su testi scientifici in lingua tedesca tradotti in italiano, selezionati tra quelli già presi in esame nello studio empirico descritto in precedenza. L'analisi è stata svolta secondo le linee guida TETASTI, sviluppate nell'ambito dello stesso studio e adattate alle caratteristiche del linguaggio scientifico tedesco-italiano.

Anche in questo studio sperimentale si è deciso di utilizzare Trados Studio come ambiente di lavoro. A tal fine, è stato scaricato e installato il plugin Opus-CAT, che consente l'implementazione del modello di TA adattato direttamente nel progetto di traduzione al momento della sua creazione. Inoltre, sono stati creati una memoria di traduzione e un database terminologico per garantire coerenza terminologica e facilitare il recupero di segmenti già tradotti, supportando in questo modo la fase di *post-editing*.

Le traduzioni generate automaticamente sono state elaborate nell'editor di testo del CAT tool e successivamente finalizzate per ottenere i file Word da esportare. Una volta ottenuti questi file, si è proceduto con l'individuazione e la classificazione degli errori, annotandoli nella sezione commenti secondo le categorie della tassonomia TETASTI. Successivamente, è stato effettuato il *post-editing* all'interno di Trados Studio, utilizzando gli stessi plug-in impiegati nello studio empirico precedente, ovvero Quality e Post-Edit Compare. Come già in precedenza illustrato, Quality permette di tracciare il tempo impiegato per svolgere l'attività di *post-editing*, e fornisce quindi informazioni utili per la valutazione del *temporal effort*. Post-Edit Compare, invece, permette di confrontare due versioni della traduzione (quella originale generata dalla TA e quella post-editata) e segnala le modifiche apportate nella fase di *post-editing*. Quest'ultimo plug-in si rivela fondamentale per la valutazione del *technical effort*.

In merito al *post-editing*, il lavoro è stato condotto replicando la metodologia dello studio empirico precedente: sono stati utilizzati due schermi, uno per visualizzare il file Word contenente l'output annotato con gli errori rilevati nella fase precedente e l'altro per procedere con la correzione direttamente su Trados Studio. Per supportare la fase di correzione sono state impiegate le seguenti risorse, già utilizzate precedentemente:

- Dizionari online bilingue: Dizionario di Tedesco – Il Sansoni Tedesco;

- Dizionari online monolingue: Duden Online per il tedesco; Treccani per l'italiano;
- Banche dati di testi paralleli: Linguee, Reverso Context;
- Motori di ricerca per la scrittura scientifica: Google Scholar;
- Strumenti per l'analisi di corpora: Sketch Engine.

Il dizionario bilingue Sansoni Tedesco è stato usato per la verifica delle corrispondenze lessicali e i dizionari monolingui Duden Online e Treccani per l'approfondimento semantico. Piattaforme come Linguee e Reverso Context hanno offerto esempi di traduzioni contestualizzate, mentre il motore di ricerca Google Scholar è stato utilizzato per controllare l'uso e le combinazioni terminologiche tipiche del linguaggio scientifico. Infine, Sketch Engine ha supportato l'analisi del corpus italiano, fornendo dati utili sulle collocazioni e sulle strutture lessicali più ricorrenti nella scrittura scientifica. La scelta di replicare la metodologia adottata nello studio empirico ha lo scopo di garantire la comparabilità e la validità dei risultati, pur lavorando in questo caso su un numero più limitato di testi a causa di vincoli temporali.

Questo approccio consente di valutare le prestazioni del sistema *fine-tuned*, verificando se le traduzioni prodotte risultano più accurate rispetto a quelle generate dal sistema generico precedentemente adottato, cioè DeepL, grazie all'addestramento su dati specificamente selezionati. Inoltre, permette di testare la robustezza e trasferibilità delle linee guida TETASTI, verificandone l'adeguatezza e l'efficacia nel contesto di un sistema di TA diverso da quello originariamente analizzato.

Nel capitolo seguente saranno presentati i risultati dell'analisi comparativa, condotta su ciascun testo tradotto con il modello *fine-tuned* e confrontato con la corrispondente traduzione generata da DeepL, al fine di evidenziare differenze qualitative e quantitative tra i due sistemi.

Testi tradotti con Deepl e il sistema di TA addestrato: un confronto

Il presente capitolo è dedicato all'analisi comparativa dei risultati ottenuti dall'annotazione degli errori e dal *post-editing* nei testi tradotti dai due sistemi di TA considerati: il modello *fine-tuned* e DeepL. L'analisi si concentra sugli output grezzi prodotti dai due sistemi, ossia sulle traduzioni automatiche prima dell'intervento umano, con l'obiettivo di esaminare in modo sistematico le differenze qualitative e quantitative tra i due output, valutando l'incidenza e la tipologia degli errori, nonché il diverso grado di intervento richiesto in fase di correzione. A tal fine, sono stati presi in esame tre testi, uno per ogni progetto di traduzione creato nello studio empirico descritto nei capitoli 5 e 6, selezionati sulla base della presenza di criticità traduttive rilevate nello studio precedente. Ciascun testo è stato tradotto tramite il modello *fine-tuned* e successivamente post-editato; i risultati così ottenuti sono stati confrontati con i rispettivi output generati da DeepL, anch'essi sottoposti a *post-editing*. I tre testi, tradotti mediante il modello adattato, sono stati raccolti in un unico progetto di traduzione su Trados Studio, al fine di agevolare l'integrazione del sistema di TA *fine-tuned* e consentire l'utilizzo di una memoria di traduzione e di un database terminologico unificati, utili per il recupero di segmenti e per garantire una gestione coerente della terminologia.

Il capitolo si articola in quattro sezioni. Le prime tre sezioni riguardano ciascuna un testo preso in esame. In ogni sezione indagine si concentra inizialmente sull'analisi comparativa degli errori individuati. I dati quantitativi sono stati raccolti in una tabella composta da tre colonne: la prima a partire da sinistra riporta le singole categorie di errore; la seconda il numero di annotazioni per ciascuna categoria nel testo tradotto dal modello *fine-tuned*; mentre nella terza rientrano il numero di annotazioni per il testo generato da DeepL.

Come per le analisi illustrate nel capitolo 6, le righe della tabella sono di colore diverso, al fine di facilitare il riconoscimento delle macrocategorie di appartenenza: il verde per *Fluency*, il giallo per *Accuracy* e il rosso per *Grammar*. Gli errori sono ordinati in ordine decrescente di frequenza sulla base del numero di annotazioni rilevate nel testo generato dal modello *fine-tuned*. Alla tabella segue un commento agli errori individuati e una riflessione comparativa supportata da esempi.

Successivamente, vengono analizzate sia le indicazioni temporali (*temporal effort*) sia quelle relative al *technical effort* emerse dalle sessioni di *post-editing*. Il *temporal effort* è stato misurato a

partire dai dati forniti dal plug-in QualityTool di Trados Studio; il *technical effort* è stato analizzato sulla base del report di Post-Edit Compare. In entrambi i casi, i dati relativi al modello *fine-tuned* sono stati messi a confronto con quelli ottenuti tramite DeepL.

Per l'analisi dell'effort tecnico è stata predisposta una tabella che riporta il PEMp (*Post-Edit Modifications percent*), ovvero la percentuale di parole del testo rimasta invariata a seguito degli interventi di *post-editing*. Come già precedentemente illustrato nel capitolo 6, questi dati vengono ripartiti in sei fasce di PEMp: 100%, 95-99%, 85-94%, 75-84%, 50-74% e *New* (0-49%). La fascia 100% indica le parole che in fase di *post-editing* sono rimaste invariate. La fascia 95-99% include i segmenti che hanno richiesto interventi minimi. Nella fascia 85-94% rientrano i segmenti che hanno subito modifiche di entità lieve, pur mantenendo una struttura complessivamente corretta. La fascia 75-84% si riferisce a traduzioni che hanno richiesto correzioni di media portata. La fascia 50-74% comprende segmenti per i quali la TA ha fornito solo un contributo parzialmente accettabile, risultando utile come base di partenza ma richiedendo un grado medio-alto di riformulazione. Infine, la categoria *New* raccoglie i segmenti con un livello di correzione compreso tra il 49% e lo 0%, cioè fino a rielaborazioni totali, evidenziando un output inadeguato. La tabella presenta per ciascuna fascia il numero di segmenti, la quota di parole e le percentuali corrispondenti, organizzati in colonne distinte e ulteriormente suddivise tra i dati del modello *fine-tuned* e quelli di DeepL.

Nella quarta e ultima sezione del capitolo viene infine proposta una riflessione complessiva derivante dall'analisi congiunta dei risultati, così da offrire una visione completa delle prestazioni dei due sistemi di TA a confronto.

8.1 Testo per il progetto n. 1 (8.454 tokens)

Corrispondente al testo n. 6 del progetto Informationen Deutsch als Fremdsprache

Brunsing, Theresa. 2016. «Landeskundliche Abbildungen in Lehrbüchern für Deutsch als Fremdsprache. Die Entwicklung des Bildeinsatzes in den Jahren 2000 bis 2010». *Informationen Deutsch als Fremdsprache* 43 (5): 494–515. <https://doi.org/doi:10.1515/infodaf-2016-0503>.

8.1.1 Analisi degli errori

Categoria di errore	Numero di annotazioni nel modello fine-tuned	Numero di annotazioni in DeepL
F/Coh/InPu	232	201
F/Sty/Disfl	203	87
A/MisTra/SemUn	98	74
F/Lex/LexCho	87	97
A/Termi/SemUn	61	60

A/Mech/CharFor	58	1
A/MisTra/WSD	52	23
G/Orth/Capit	50	19
A/DNT	31	17
A/Omis	24	18
A/Untra	21	14
F/Sty/Reg	14	19
G/WF/Agree	13	11
A/MisTra/Part	12	13
G/Orth/Punct	11	11
F/Lex/Fue	10	10
A/MisTra/WroPr	9	5
G/Orth/Spel	7	6
F/Coh/CoRef	7	2
A/ForLa	6	17
F/Coh/Cit	6	8
A/MisOrd/WO	4	13
A/Add	4	8
A/Mech/Punct	4	3
G/Orth/WhiSpa	3	13
A/MisTra/POS	3	5
F/De/Layo	3	3
F/SLP/Nomin	2	21
A/MWE/Ell	2	2
A/MisExpl	2	2
A/MisOrd/MW	2	-
A/MWE/Komp	2	-
F/Sty/Idiom	2	-
F/SLP/ForImp	1	3
A/SouEr	1	1
A/MisTra/Gend	1	1
A/Termi/DNT	-	12

Tabella 11: Analisi quantitativa comparativa degli errori per il testo del modello *fine-tuned* n. 1

Nel testo tradotto mediante il sistema *fine-tuned* sono stati rilevati 565 errori di *Fluency*, 399 di *Accuracy* e 84 di *Grammar*, per un totale di 1.048 annotazioni. Circa la metà del totale degli errori ricade nella macrocategoria *Fluency*, principalmente a causa della categoria F/Coh/InPu, che con 232 occorrenze risulta la più rappresentata. Tale categoria riguarda l'uso incoerente di segni ortografici, quali l'apostrofo e le virgolette (cfr. esempio n. 1). Invece, nel testo tradotto con il sistema di TA generico DeepL si riscontrano 451 errori di *Fluency*, 289 di *Accuracy* e 60 di *Grammar*. Il totale degli errori annotati ammonta a 800. Anche in questo caso la quota prevalente degli errori rientra nella macrocategoria *Fluency*, che supera la metà del totale, e la categoria F/Coh/InPu, con 201 occorrenze, è la più rappresentata. Nel complesso il sistema *fine-tuned*

presenta quindi un numero più elevato di errori sia nelle singole macrocategorie che nel conteggio totale.

*Esempio n. 1 (ProgFT_Testo1 – F/Coh/InPu)*⁷⁸

TAFT: [...] la ricezione e l'interpretazione [...]

PEFT: [...] la ricezione e l'interpretazione [...]

In ordine decrescente per numero di occorrenze, nel testo del modello *fine-tuned* seguono le seguenti categorie di errori: F/Sty/Disfl con 203 errori e A/MisTra/SemUn con 98. Anche per il testo tradotto con DeepL, queste categorie registrano un maggior numero di risultati: 87 occorrenze per F/Sty/Disfl e 74 per A/MisTra/SemUn. Si riscontrano quindi dei valori più elevati nel testo tradotto dal modello *fine-tuned*, dove in particolare la categoria F/Sty/Disfl, che riguarda nello specifico le difficoltà di scorrevolezza della lingua di arrivo, risulta particolarmente problematica. Invece, gli errori di A/MisTra/SemUn segnalano la mancata corrispondenza semantica tra testo di partenza e testo di arrivo. Di seguito viene riportato un esempio per entrambe le categorie tratto dal testo del modello *fine-tuned*.

Esempio n. 2 (ProgFT_Testo1 – F/Sty/Disfl + A/MisTra/SemUn)

TO: Durch das obige Zitat wird in prägnanter Weise die Dominanz des Gesichtssinns gegenüber allen anderen Sinnen sichtbar.

TAFT: Attraverso la citazione di cui sopra si vede, in maniera concisa, la dominanza del senso del volto nei confronti di tutti gli altri sensi.

PEFT: La citazione di cui sopra illustra, in maniera concisa, la dominanza del senso della vista su tutti gli altri sensi.

L'esempio mostra la co-occorrenza di un errore di scorrevolezza (F/Sty/Disfl) e di un errore di traduzione semanticamente errata (A/MisTra/SemUn). L'espressione "senso del volto" costituisce infatti un errore semantico derivante da un'interpretazione letterale del lessema tedesco "Gesichtssinn", il cui significato corretto è "senso della vista". Tale scelta compromette la correttezza informativa del testo di arrivo, configurandosi come errore di A/MisTra/SemUn. Inoltre, la struttura sintattica della frase risulta poco naturale, in particolare a causa dei costrutti "attraverso

⁷⁸ Nel presente capitolo, per agevolare la distinzione tra gli esempi tratti dal testo del modello *fine-tuned* e quelli tratti dal testo di DeepL, si adottano le seguenti sigle:

TAFT = traduzione automatica prodotta dal modello *fine-tuned*;

PEFT = *post-editing* della traduzione prodotta dal modello *fine-tuned*;

TA = traduzione automatica prodotta da DeepL;

PE = *post-editing* della traduzione prodotta da DeepL.

la citazione di cui sopra si vede” e “nei confronti di tutti gli altri sensi”. Per questo motivo è necessario un intervento di riformulazione, classificato come errore F/Sty/Disfl.

Un’ulteriore categoria che registra un numero di errori particolarmente consistente in entrambi i testi è F/Lex/LexCho. Questa riguarda l’impiego di espressioni non idonee all’ambito specialistico e al genere testuale considerato. Per tale categoria si registrano 87 errori nel testo tradotto dal modello *fine-tuned* e 97 nel testo generato da DeepL, che mostra quindi una maggiore frequenza. Di seguito vengono presentati due esempi: l’esempio n. 3 per il testo generato dal modello *fine-tuned* e l’esempio n. 4 per il testo di DeepL.

Esempio n. 3 (ProgFT_Testo1 – F/Lex/LexCho)

TO: Weidenmann teilt diese Funktion in vier untergeordnete Kategorien ein: Aktivierungs-, Konstruktions-, Fokus- und Ersatzfunktion.

TAFT: Weidenmann classifica questa funzione in quattro categorie subordinate: funzione di attivazione, costruzione, messa a fuoco e sostituzione.

PEFT: Weidenmann classifica questa funzione in quattro categorie subordinate: funzione di attivazione, costruzione, focalizzazione e sostituzione. [...]

Esempio n. 4 (Prog1_Testo6 – F/Lex/LexCho)

TO: Häufige Ziele von Informationsbildern sind: [...]

TA: I bersagli frequenti delle immagini informative sono [...]

PE: Gli obiettivi frequenti delle immagini informative sono [...]

In entrambi gli esempi emerge la difficoltà dei sistemi di TA nel selezionare il lessema più appropriato al linguaggio scientifico. Nello specifico, le scelte lessicali, pur grammaticalmente corrette, risultano stilisticamente inadeguate al genere testuale e all’ambito specialistico trattati, compromettendo la naturalezza e la coerenza del testo e del linguaggio.

Nel testo prodotto dal sistema *fine-tuned* si registrano 61 errori di A/Termi/SemUn, relativi alla traduzione errata di termini, mentre nel testo tradotto da DeepL ne sono stati individuati 60. Per la categoria A/Termi/DNT, impiegata per indicare termini che non necessitano di traduzione, non si riscontrano errori nel testo del modello *fine-tuned*, a differenza di quello di DeepL, dove sono stati individuati 12 errori. Gli esempi seguenti illustrano alcuni casi significativi di errori terminologici nei due sistemi.

Esempio n. 5 (ProgFT_Testo1 – A/Termi/SemUn)

TO: Da in der Analyse nur landeskundliche Abbildungen berücksichtigt wurden, [...]

TAFT: Dato che nell’analisi sono state prese in considerazione solo le illustrazioni nazionali, [...]

PEFT: Dato che nell’analisi sono state prese in considerazione solo le illustrazioni riguardanti la cultura e la civiltà del Paese, [...]

Esempio n. 6 (Prog1_Testo6 – A/Termi/SemUn)

TO: Da in der Analyse nur landeskundliche Abbildungen berücksichtigt wurden, [...]

TA: Poiché nell’analisi sono state prese in considerazione solo le illustrazioni degli studi regionali, [...]

PE: Poiché nell’analisi sono state prese in considerazione solo le illustrazioni riguardanti la cultura e la civiltà del Paese, [...]

Negli esempi n. 5 e n. 6, entrambi i sistemi traducono in modo errato il termine “landeskundliche”, generando errori di A/Termi/SemUn. Il modello *fine-tuned* propone “nazionali”, mentre DeepL “degli studi regionali”, entrambe soluzioni che non restituiscono il significato specialistico del termine. Il concetto di *Landeskunde* è sempre stato e continua a essere mutevole e, per questo, difficile da circoscrivere (Penning 1995). Risulta tuttavia particolarmente esaustivo quanto affermato da Reinbothe (1997) in riferimento all’oggetto di studio di tale disciplina applicata alla Germania: “Gegenstand der Landeskunde sind die geographischen, ökonomischen, sozialen, politischen und kulturellen Verhältnisse in Deutschland, deren Geschichte und Begriffe”. Si tratta dunque di un termine complesso, che richiama aspetti della geografia, dell’economia e più in generale delle scienze sociali e culturali, per il quale non esiste un equivalente diretto in italiano. In fase di *post-editing* si è deciso di impiegare la formulazione “riguardanti la cultura e la civiltà del Paese”, che rappresenta la soluzione più vicina al significato del termine originale. Dunque, in entrambi gli esempi n. 5 e n. 6, il significato specialistico del termine non è stato correttamente trasferito nel testo di arrivo, evidenziando come i due sistemi riscontrino difficoltà nella resa di termini specifici del contesto culturale e in particolare nell’ambito didattico.

Esempio n. 7 (Prog1_Testo6 – A/Termi/DNT)

PE: In der Literatur finden sich etliche Definitionen dessen, was Visual Literacy ist [...]

TA: In letteratura esistono diverse definizioni di visualità alfabetizzazione [...]

PE: In letteratura esistono diverse definizioni di visual literacy [...]

Nell’esempio n.7 l’errore riguarda la traduzione di un termine inglese ormai consolidato nel linguaggio scientifico italiano, ciò giustifica la scelta di mantenere il termine in lingua inglese (“visual literacy”). DeepL, invece, mostra una gestione problematica dei termini in lingua straniera: non riconosce quando l’uso è consolidato nella lingua di arrivo e traduce letteralmente anche

termini ormai stabilizzati, producendo traduzioni semanticamente imprecise e incorrendo in errori della categoria A/Termi/DNT.

Altre categorie che registrano un numero significativo di occorrenze nel testo del modello *fine-tuned* sono A/Mech/CharFor (con 58 occorrenze) riguardante gli errori nel trasferimento della formattazione e A/MisTra/WSD (con 52 occorrenze) per gli errori di scelta semantica in riferimento a parole polisemiche. Tali categorie presentano un numero considerevolmente inferiore nel testo tradotto da DeepL: una sola occorrenza per la categoria A/Mech/CharFor e 23 per la categoria A/MisTra/WSD. Di seguito vengono riportati degli esempi tratti dal testo del modello adattato.

Esempio n. 8 (ProgFT_Testo1 – A/Mech/CharFor)

TO: 2. Bilder und ihre Wirkung

TAFT: 2. Le immagini e il loro effetto

PEFT: 2. Le immagini e il loro effetto

L'esempio n. 8 illustra un errore di A/Mech/CharFor: nella proposta della TAFT, infatti, solo alcune parole risultano in grassetto, mentre nel testo originale l'intero segmento "2. Bilder und ihre Wirkung" presenta una formattazione uniforme. Tale problema è riconducibile a una gestione non coerente dei segmenti da parte del sistema *fine-tuned*, che tende frequentemente a non riprodurre correttamente le caratteristiche tipografiche del testo di origine (come stile, grandezza, uso del corsivo o del grassetto), compromettendo la coerenza visiva e strutturale del testo.

Esempio n. 9 (ProgFT_Testo1 – A/MisTra/WSD)

TO: Eine weitere Funktion ist die Wissensvermittlung, [...]

TAFT: Un'altra funzione è la mediazione del sapere [...]

PEFT: Un'altra funzione è la trasmissione del sapere [...]

L'esempio n. 9, invece, presenta un errore di A/MisTra/WSD. Il lessema "Wissensvermittlung" è stato tradotto con "mediazione del sapere", anziché "trasmissione del sapere", che rappresenta la resa più appropriata nel contesto specialistico scientifico-didattico trattato nel testo. L'errore deriva da una scelta non adeguata fra le possibili accezioni del lessema, generando quindi una discrepanza semantica pur in presenza di una frase grammaticalmente corretta.

In riferimento alla macrocategoria *Grammar*, si osservano differenze significative nella distribuzione degli errori tra i due sistemi di TA. Questa categoria comprende errori che interessano diversi aspetti grammaticali e ortografici e, in generale, la correttezza formale della lingua di arrivo. Le sottocategorie individuate sono: accordo errato tra parole (G/WF/Agree); errori di punteggiatura

(G/Orth/Punct) e di spaziatura (G/Orth/WhiSpa); errori ortografici (G/Orth/Spel), errori relativi all'uso della lettera maiuscola (G/Orth/Capit).

Nel testo prodotto dal sistema *fine-tuned*, gli errori risultano complessivamente più numerosi rispetto a DeepL. La distribuzione rilevata è la seguente: G/Orth/Capit con 50 errori, G/WF/Agree 13 errori, G/Orth/Punct con 11 errori, G/Orth/Spel con 7 errori, G/Orth/WhiSpa con 3 errori. Nel testo generato da DeepL, invece, gli errori si distribuiscono nel seguente modo: G/Orth/Capit con 19 errori, G/WF/Agree e G/Orth/Punct con 11 errori ciascuna, G/Orth/Spel con 6 errori e G/Orth/WhiSpa con 13 errori. Dunque, il sistema *fine-tuned* mostra una maggiore incidenza di errori di accordo tra parole (G/WF/Agree), errori ortografici (G/Orth/Spel) ed errori relativi all'uso della lettera maiuscola (G/Orth/Capit), dove per quest'ultima categoria la differenza con DeepL è particolarmente marcata (cfr. esempio n. 10). Il sistema DeepL, invece, registra un numero più elevato di errori di spaziatura (G/Orth/WhiSpa, cfr. esempio n. 11); mentre per gli errori di punteggiatura (G/Orth/Punct) entrambi i sistemi riportano lo stesso numero di occorrenze.

Esempio n. 10 (ProgFT_Testo1 – G/Orth/Capit)

TO: [...] und zusammen die Visual Literacy fördern [...]

TAFT: [...] e insieme promuovono la Visual Literacy [...]

PEFT: [...] e insieme promuovono la visual literacy [...]

L'errore presentato nell'esempio sopra riportato deriva dal trasferimento diretto della convenzione ortografica tedesca riguardante l'uso della lettera maiuscola. Tale interferenza produce nel testo di arrivo l'uso improprio della lettera maiuscola, a scapito della correttezza formale.

Esempio n. 11 (Prog1_Testo6 – G/Orth/WhiSpa)

TA: [...] due libri di testo, *Optimal* e *Schritte*, contengono [...]

PE: [...] due libri di testo, *Optimal* e *Schritte*, contengono [...]

Nell'esempio n. 11, l'errore consiste nella presenza di uno spazio superfluo prima della virgola, rappresentando il mancato rispetto delle norme tipografiche. Pur non influenzando il significato del testo, tale errore ne compromette la correttezza formale e la leggibilità.

Proseguendo nell'analisi si riscontrano le seguenti categorie: A/DNT (31 errori nel testo del modello *fine-tuned* e 17 in quello di DeepL); A/Omis (24 errori nel testo del sistema adattato e 18 in quello di DeepL); A/Untra (21 errori nel testo del modello *fine-tuned* e 14 in quello di DeepL). Ciò mostra chiaramente la tendenza del sistema *fine-tuned* a registrare più errori rispetto a DeepL nelle categorie considerate. Gli errori di A/Omis riguardano l'omissione di un segmento e possono interessare singole parole, più parole o intere frasi. Invece, gli errori di A/DNT riguardano i

segmenti la cui traduzione non è necessaria. Nel testo preso in esame, tali errori sono specificamente legati alla tematica dell'articolo, che concerne l'analisi di alcuni libri di testo. In particolare, le problematiche riguardano la traduzione impropria dei titoli dei manuali, dovuta all'incapacità dei sistemi di TA di riconoscerli come tali (cfr. esempio n. 12 e 13). Tale fenomeno è quantitativamente più consistente nel testo generato dal modello *fine-tuned*.

Esempio n. 12 (ProgFT_Testo1 – A/DNT)

TO: Schritte plus geht noch weiter, [...]

TAFT: Passi più va oltre, [...]

PEFT: Schritte plus va oltre, [...]

Esempio n. 13 (Prog1_Testo6 – A/DNT)

TO: Schritte plus geht noch weiter, [...]

TA: Passi plus si spinge oltre, [...]

PE: Schritte plus si spinge oltre [...]

Gli errori di A/Untra invece segnalano i segmenti per i quali il sistema di TA non ha fornito alcuna traduzione. Nel testo generato dal modello *fine-tuned*, ciò riguarda singoli lessemi (cfr. esempio n. 14), mentre in DeepL gli errori interessano frasi intere (cfr. esempio n. 15).

Esempio n. 14 (ProgFT_Testo1 – A/Untra)

TO: Die Ergebnisse aus der Forschung finden bisher jedoch wenig Beachtung im DaF-Unterricht (vgl. Bachtsevanidis 2012).

TAFT: I risultati della ricerca, tuttavia, sono stati finora scarsamente presi in considerazione nella didattica del DaF (cfr. Bachtsevanidis 2012).

PEFT: I risultati della ricerca, tuttavia, sono stati finora scarsamente presi in considerazione nella didattica del tedesco come lingua straniera (cfr. Bachtsevanidis 2012).

Esempio n. 15 (Prog1_Testo6 – A/Untra)

TO: Den ersten interkulturellen Vergleich enthält Themen aktuell (2004), jedoch im gesamten Lehrbuch nur an zwei Stellen (vgl. Perlmann-Balme et al. 2004: 50, 79).

TA: Den ersten interkulturellen Vergleich enthält Themen aktuell (2004), jedoch im gesamten Lehrbuch nur an zwei Stellen (vgl. Perlmann-Balme et al. 2004: 50, 79).

PE: Il primo confronto interculturale è presente in Themen aktuell (2004), tuttavia in tutto il manuale esso compare soltanto in due punti (cfr. Perlmann-Balme et al. 2004: 50, 79).

Gli esempi riportati evidenziano differenze nel comportamento dei due sistemi, che possono essere spiegate dal diverso approccio con cui questi trattano i segmenti. Il modello *fine-tuned* tende a non tradurre singoli lessemi, intervenendo solo dove riesce a fornire una traduzione affidabile, mentre lascia invariati altri elementi del segmento. DeepL, invece, affronta i segmenti in maniera più globale e, quando incontra difficoltà interpretative su frasi complesse o poco contestualizzate, può omettere la traduzione di intere frasi o porzioni di testo, generando quindi lacune più estese nel testo di arrivo.

Per quanto attiene lo stile del linguaggio scientifico, si segnalano 14 errori di registro inadeguato (F/Sty/Reg) per il testo generato dal sistema *fine-tuned* e 19 errori nel testo prodotto da DeepL. Ciò indica che il modello adattato gestisce meglio l'uso di un registro adeguato al linguaggio scientifico rispetto al sistema di TA generico (cfr. esempi n. 16 e n. 17).

Esempio n. 16 (ProgFT_Testo1 – F/Sty/Reg)

TO: [...] um die Seiten ansprechend zu gestalten.

TAFT: [...] per rendere le pagine accattivanti.

Esempio n. 17 (Prog3_Testo1 – F/Sty/Reg)

TO: [...] um die Seiten ansprechend zu gestalten.

TA: [...] per rendere le pagine più attraenti

PE: [...] per rendere le pagine accattivanti.

Negli esempi riportati si osserva una differenza nell'uso del registro linguistico tra i due sistemi di TA. Nel caso del modello *fine-tuned* (esempio n. 16), il segmento tradotto “per rendere le pagine accattivanti” risulta adeguato al registro del linguaggio scientifico, mantenendo una neutralità lessicale compatibile con il genere testuale trattato. Al contrario, nel testo generato da DeepL (esempio n. 17), la scelta lessicale “più attraenti” introduce un registro colloquiale e incoerente con lo stile dell'articolo scientifico. Inoltre, nell'esempio n. 17 si osserva anche un errore di tipo A/Add: l'aggiunta di “più” aggiunge all'aggettivo “attraenti” il grado comparativo di maggioranza non presente nel segmento di partenza, alterando parzialmente il significato originale.

Per quanto riguarda il lessico del linguaggio scientifico, sia nel testo prodotto dal sistema *fine-tuned* che in quello generato da DeepL, sono stati riscontrati 10 errori di F/Lex/Fue, riferibili alla traduzione errata delle *Fügungen*. È interessante notare che, in alcuni casi, i due testi presentano errori per gli stessi segmenti di testo, come mostrato negli esempi n. 18 e n. 19, mentre in altri casi riguardano *Fügungen* differenti, a testimonianza di una difficoltà generalizzata da parte di entrambi

i sistemi nel rendere correttamente queste combinazioni tipiche del linguaggio scientifico quotidiano, compromettendone quindi l'autenticità e la naturalezza.

Esempio n. 18 (ProgFT_Testo1 – F/Lex/Fue)

TO: Hier steht nicht allein das Bild im Zentrum, sondern [...]

TAFT: Qui non c'è solo l'immagine al centro, ma [...]

PEFT: Qui l'attenzione non si concentra solo sull'immagine, ma [...]

Esempio n. 19 (Prog3_Testo1 – F/Lex/Fue)

TO: Hier steht nicht allein das Bild im Zentrum, sondern [...]

TA: Qui non è solo l'immagine ad essere al centro della scena, ma [...]

PE: Qui l'attenzione non si concentra solo sull'immagine, ma [...]

Negli esempi riportati, entrambi i sistemi producono una resa letterale della *Fügung* tedesca “im Zentrum stehen”, senza coglierne il significato e la funzione all'interno del testo di partenza.

In riferimento ai fenomeni linguistici del linguaggio scientifico, nel testo prodotto con il modello *fine-tuned* si registrano le seguenti occorrenze: 2 errori di F/SLP/Nomin, riguardanti il mancato impiego di nominalizzazioni e un errore di F/SLP/ForImp, connesso all'impiego delle forme personali in sostituzione di quelle impersonali. Nel testo prodotto da DeepL si riscontrano invece 21 errori di F/SLP/Nomin e 3 errori per la categoria F/SLP/ForImp. In riferimento alla categoria di F/SLP/Nomin, i valori si differenziano in modo significativo: DeepL tende a privilegiare strutture verbali esplicite al posto delle nominalizzazioni, mentre le proposte del modello *fine-tuned* si dimostrano maggiormente conformi alle convenzioni stilistiche del genere testuale.

Esempio n. 20 (ProgFT_Testo1 – F/SLP/Nomin)

TO: Durch die gezielte Übung der Wahrnehmung kultureller Unterschiede [...]

TAFT: Attraverso l'esercizio mirato della percezione delle differenze culturali, [...]

Esempio n. 21 (Prog3_Testo1 – F/SLP/Nomin)

TO: Durch die gezielte Übung der Wahrnehmung kultureller Unterschiede [...]

TA: Esercitando in modo specifico la percezione delle differenze culturali, [...]

PE: Attraverso l'esercizio mirato della percezione delle differenze culturali, [...]

Negli esempi n. 20 e n. 21, si osserva come DeepL trasformi la nominalizzazione tedesca “durch die gezielte Übung der Wahrnehmung” in una forma verbale (“Esercitando in modo specifico...”), introducendo una prospettiva agentiva che rende esplicita l'azione, a scapito dell'astrattezza

concettuale tipica dello stile nominale scientifico. Il modello *fine-tuned*, invece, mantiene la struttura nominale, preservando un'impostazione più impersonale, maggiormente conforme alle convenzioni stilistiche del testo scientifico.

Anche la categoria F/SLP/ForImp registra più occorrenze di errore nel testo tradotto da DeepL: 3 errori a fronte di uno nel testo del modello *fine-tuned*. La differenza è contenuta, ma conferma l'andamento osservato finora in riferimento agli errori di resa dello stile e dei fenomeni del linguaggio scientifico.

Esempio n. 22 (ProgFT_Testo1 – F/SLP/Nomin)

TO: Zur Beantwortung wird untersucht, ob [...]

TAFT: Per rispondere si indaga se [...]

Esempio n. 23 (Prog3_Testo1 – F/SLP/Nomin)

TO: Zur Beantwortung wird untersucht, ob [...]

TA: Per rispondere a questa domanda, analizzeremo se [...]

PE: Per rispondere a questa domanda, si analizzerà se [...]

Nell'esempio n. 23, si osserva come DeepL preferisca la forma personale (“analizzeremo”) al posto di quella impersonale (“si analizzerà”), riducendo il grado di oggettività tipico dello stile scientifico. Il modello *fine-tuned*, invece, mantiene la costruzione impersonale, più conforme alle convenzioni del linguaggio scientifico (cfr. esempio n. 22). La differenza del tempo verbale (presente nel primo caso e futuro nel secondo) non è stata considerata un errore, in quanto risulta coerente con l'uso verbale adottato nel resto dei due testi.

Proseguendo nell'analisi degli errori, si riscontrano le seguenti categorie di *Accuracy*: A/MisTra/Part, A/MisTra/WroPr e A/ForLa. Gli errori di A/MisTra/Part riguardano il trasferimento parziale o impreciso del significato, e più nello specifico interessano il trasferimento errato di modi verbali o del genere e numero di un lessema (cfr. esempi n. 67 e n. 68 in 8.2.1). È importante sottolineare che tali errori non compromettono la correttezza grammaticale della lingua di arrivo: la criticità riguarda invece il piano semantico, poiché il trasferimento incompleto o impreciso del contenuto informativo produce un significato parzialmente divergente rispetto al testo di partenza. Per tale categoria si registrano 12 occorrenze nel testo prodotto dal modello *fine-tuned* e 13 nel testo di DeepL. La differenza risulta minima, segnalando pari difficoltà da parte dei due sistemi. La categoria A/MisTra/WroPr, riguardante gli errori nell'uso di preposizioni, sintagmi preposizionali o congiunzioni, registra 9 occorrenze nel testo tradotto dal sistema adattato e 5 nel testo generato da

DeepL, indicando quindi maggiori criticità da parte del modello *fine-tuned*. Di seguito viene riportato un esempio tratto dal testo del sistema adattato.

Esempio n. 24 (ProgFT_Testo1 – A/MisTra/WroPr)

TO: Die Bildproduktion besteht bei Bachtsevanidis in einem produktiven Umgang mit Bildern, [...]

TAFT: La produzione di immagini di Bachtsevanidis consiste in una gestione produttiva delle immagini [...]

PEFT: Per Bachtsevanidis, la produzione di immagini consiste in una gestione produttiva delle immagini [...]

In questo caso, l'impiego della preposizione errata compromette il significato della frase, ma non la correttezza grammaticale, alterando il rapporto sintattico fra i costituenti e realizzando un'interpretazione non prevista dal testo di partenza. Inoltre, il segmento “bei Bachtsevanidis” è interessato da un errore di F/Sty/Disfl: l'intervento di correzione ha comportato una riorganizzazione dell'ordine degli elementi nella frase per garantire maggiore scorrevolezza e naturalezza della lingua di arrivo.

La categoria A/ForLa invece riguarda l'uso improprio di una lingua straniera, in particolare l'inglese, in sostituzione della lingua di arrivo richiesta. Per tale categoria si registrano 6 errori nel testo tradotto dal modello *fine-tuned* e 17 nel testo generato da DeepL, evidenziando un maggiore ricorso all'inglese come lingua *pivot* da parte del sistema generico. Di seguito vengono riportati due esempi: l'esempio n. 25 per il testo di DeepL e l'esempio n. 26 per il testo del modello adattato.

Esempio n. 25 (Prog1_Testo6 – A/ForLa)

TO: Daneben gibt es in Schritte plus neue Aufgaben, die [...]

TA: Ci sono anche nuovi compiti in Steps plus che [...]

PE: Ci sono anche nuovi compiti in Schritte plus che [...]

Nel testo tradotto da DeepL, questa categoria di errore riguarda principalmente i titoli dei libri di testo, presi in esame nell'articolo, che vengono erroneamente tradotti in inglese invece di essere mantenuti nella forma originale tedesca, come mostra l'esempio summenzionato. Un fenomeno analogo si osserva anche nel testo prodotto dal modello *fine-tuned*; in quest'ultimo, tuttavia, si registrano anche casi in cui singoli lessemi vengono tradotti in inglese all'interno di frasi altrimenti corrette in italiano (cfr. esempio n. 26).

Esempio n. 26 (ProgFT_Testo1 – A/ForLa)

TO: [...] die den Lernenden als Orientierung dienen [...]

TAFT: [...] che servono come orientation per gli apprendenti [...]

PEFT: [...] che servono come orientamento per gli apprendenti [...]

In entrambi i casi, questo fenomeno può essere spiegato alla luce del funzionamento dei modelli di TA neurale. Quando il sistema non riconosce un equivalente consolidato nella lingua di arrivo, l'inglese può agire implicitamente come lingua *pivot*, portando il modello a selezionare lessemi inglesi come soluzione intermedia. Tale comportamento può produrre discrepanze o imprecisioni semantiche nella resa finale del testo.

Altri errori di *Fluency* che si riscontrano riguardano le seguenti categorie: F/Coh/CoRef che riguarda la mancata corrispondenza tra referenti citati in frasi diverse (7 errori nel testo del modello *fine-tuned* e 2 nel testo di DeepL); F/Coh/Cit che segnala i casi di strutturazione incoerente dei riferimenti bibliografici (6 errori nel testo del sistema adattato e 8 nel testo di DeepL); e F/De/Layo che concerne gli errori di strutturazione dei paragrafi e del testo di arrivo (3 errori nel testo del modello *fine-tuned* e nel testo di DeepL). In riferimento alla categoria F/Coh/CoRef si individuano più errori nel testo tradotto dal modello *fine-tuned*, a indicare una maggiore difficoltà da parte di quest'ultimo nel garantire la coerenza testuale (cfr. esempio n. 27).

Esempio n. 27 (ProgFT_Testo1 – F/Coh/CoRef)

TO: Sie können Schemata oder mentale Modelle darstellen, [...]

TAFT: Essi possono rappresentare schemi o modelli mentali [...]

PEFT: Queste possono rappresentare schemi o modelli mentali [...]

Nell'esempio n. 27, l'errore riguarda la mancata corrispondenza tra il pronome soggetto e il referente originale: il pronome "sie" fa riferimento alla parola "immagini", citata nella frase precedente. La forma corretta, come mostrato nel *post-editing*, è "queste", che garantisce la coerenza anaforica e la corretta interpretazione del testo.

La categoria F/Coh/Cit è una delle categorie più significative per la traduzione del linguaggio scientifico. I cinque errori presenti nel testo del modello *fine-tuned* ricorrono anche in quello di DeepL e riguardano lo stesso fenomeno, come mostrato dall'esempio n. 28.

Esempio n. 28 (ProgFT_Testo1 – F/Coh/Cit)

TO: [...] (ebd.: 35f.).

TAFT: [...] (ibid.: 35f.).

PEFT: [...] (ibid.: 35-36).

In questo caso, l'errore di F/Coh/Cit riguarda l'abbreviazione "f." (indicante la pagina successiva), non conforme alle convenzioni italiane per i riferimenti bibliografici. Il *post-editing* corregge l'indicazione delle pagine nel rispetto dello stile italiano e delle norme citazionali richieste dalla rivista⁷⁹. Tuttavia, il sistema generico presenta ulteriori difficoltà nella gestione dei riferimenti bibliografici, come evidenziato dall'esempio seguente.

Esempio n. 29 (Prog1_Testo6 – F/Coh/Cit)

TO: [...] (ebd.).

TA: [...] (ibidem).

PE: [...] (ibid.).

Nell'esempio n. 29 emerge una variazione impropria: DeepL utilizza qui "ibidem", mentre in altri casi analoghi dello stesso testo aveva impiegato precedentemente "ibid.". Questa alternanza genera una disomogeneità nelle convenzioni citazionali, compromettendo l'uniformità formale del testo e riflettendo una gestione meno sistematica delle abbreviazioni bibliografiche rispetto al modello *fine-tuned*, che invece mantiene una resa più stabile e coerente.

Tra le categorie che registrano meno errori nel testo tradotto dal modello *fine-tuned* si individuano le seguenti: A/MisOrd/WO (errori riconducibili alle variazioni dell'ordine degli elementi), A/Add (aggiunta impropria di segmenti) e A/Mech/Punct (trasferimento errato della punteggiatura) con 4 occorrenze ciascuna e A/MisTra/POS (impiego della categoria lessicale errata) con 3 occorrenze. Nel testo generato da DeepL, invece, gli errori relativi a queste categorie risultano più numerosi: A/MisOrd/WO presenta 13 occorrenze, A/Add 7 e A/MisTra/POS 5. L'unica eccezione è rappresentata da A/Mech/Punct, che con 3 occorrenze risulta leggermente inferiore rispetto al modello *fine-tuned*. Tale categoria, riguardante gli errori di trasferimento della punteggiatura, comporta interventi minori dal punto di vista traduttivo, ma che possono comunque compromettere la leggibilità e l'accuratezza del testo di arrivo.

Pertanto, i dati riportati indicano una gestione più problematica da parte di DeepL soprattutto nell'ordine degli elementi all'interno della frase (A/MisOrd/WO), determinando variazioni di significato e pregiudicando la corretta trasmissione del messaggio del testo di partenza, ma non la correttezza grammaticale della lingua di arrivo (cfr. esempio n. 30).

Esempio n. 30 (Prog1_Testo6 – A/MisOrd/WO)

TO: Bilder, die die Phantasie ansprechen, wirken gleichzeitig motivierend, [...]

TA: Anche le immagini che richiamano l'immaginazione hanno un effetto motivante, [...]

⁷⁹ Per maggiori informazioni si veda <https://share.google/b4vVHnmAHoNZvNOWK> (ultima consultazione: 20/12/2025)

PE: Le immagini che richiamano l'immaginazione hanno allo stesso tempo un effetto motivante, [...]

Nell'esempio n. 30 si osserva inoltre un errore di resa semantica (A/MisTra/SemUn) nella traduzione del termine "gleichzeitig", reso con "anche" anziché con un'espressione che indichi simultaneità. Tale resa comporta non solo la perdita del valore di contemporaneità presente nel testo di partenza, ma introduce un rapporto additivo estraneo all'originale, con un conseguente mutamento dell'interpretazione dell'enunciato.

DeepL presenta anche più casi di aggiunta di segmenti (A/Add) e d'impiego della categoria lessicale (*Wortart*) non corrispondente a quella del testo di partenza (A/MisTra/POS). Per quest'ultima categoria viene riportato di seguito un esempio tratto da testo tradotto da DeepL.

Esempio n. 31 (Prog1_Testo6 – A/MisTra/POS)

TO: Welche [...] Funktionen erfüllen die landeskundlichen Abbildungen [...]?

TA: Quali funzioni [...] svolgono gli studi regionali illustrati [...]?

PE: Quali funzioni [...] svolgono le illustrazioni riguardanti la cultura e la civiltà [...]?

Nell'esempio summenzionato, si verificano due casi di A/MisTra/POS, riguardanti il sostantivo "Abbildungen" e l'aggettivo "landeskundlichen", resi nella proposta di traduzione automatica rispettivamente come participio passato con funzione aggettivale ("illustrati") e come sintagma formato da sostantivo + aggettivo ("studi regionali"). Questa sostituzione della categoria lessicale altera il significato complessivo del segmento. In fase di *post-editing* è stata ripristinata la corretta categoria lessicale dei lessemi, intervenendo al contempo sulla resa dell'aggettivo "landeskundlichen", lessema problematico ricorrente nel testo e già discusso nell'esempio n. 6.

Le categorie A/MWE/Ell e A/MisExpl registrano lo stesso numero di errore in entrambi i testi tradotti automaticamente: due occorrenze per ciascuna categoria. La categoria A/MWE/Ell riguarda gli aspetti problematici legati alla traduzione di composti con ellissi di elementi ripetuti; mentre gli errori della categoria A/MisExpl sono riferibili all'inadeguatezza dei contenuti per il destinatario del testo di arrivo. Di seguito vengono riportati due esempi per quest'ultima categoria.

Esempio n. 32 (ProgFT_Testo1 – A/MisExpl)

TO: Der Satz »Herzlich willkommen!« soll [...] ausgesprochen werden.

TAFT: La frase "Benvenutissimo!" deve essere pronunciata [...]

PEFT: La frase "Herzlich willkommen!" (Benvenuto!) deve essere pronunciata [...]

Esempio n. 33 (Prog1_Testo6 – A/MisExpl)

TO: Der Satz »Herzlich willkommen!« soll [...] ausgesprochen werden.

TA: La frase "Benvenuto!" deve essere pronunciata [...]

PE: La frase "Herzlich willkommen!" (Benvenuto!) deve essere pronunciata [...]

Negli esempi n. 32 e n. 33, l'espressione "Herzlich willkommen!" fa riferimento a un esercizio presente in uno dei manuali analizzati nel testo. In fase di *post-editing*, l'inserimento dell'espressione tedesca tra virgolette e della relativa traduzione italiana tra parentesi consente di preservare la fedeltà al testo di partenza e, al contempo, di adattare il contenuto al destinatario italiano, garantendone la comprensione. Si osserva, inoltre, in entrambi gli esempi un errore di F/Coh/InPu, relativo all'uso incoerente delle virgolette. Infine, nell'esempio n. 32, tratto dal testo del modello *fine-tuned*, si riscontra anche un errore di F/Sty/Reg, dovuto all'impiego del lessema "Benvenutissimo", non adeguato al registro del linguaggio scientifico.

In seguito nell'analisi della tabella 13, si può constatare che le categorie A/MisOrd/MW, A/MWE/Komp e F/Sty/Idiom presentano ciascuna 2 occorrenze nel testo prodotto dal modello *fine-tuned* e nessuna nel testo tradotto da DeepL. La categoria A/MisOrd/MW segnala i casi di scorretta attribuzione di elementi all'interno della frase, che compromettono la corretta trasmissione del messaggio del segmento di partenza ma non la correttezza grammaticale della lingua di arrivo (cfr. esempio n. 71 in 8.2.1). Invece, gli errori della categoria A/MWE/Komp derivano da una traduzione problematica di aggettivi composti, che non tiene conto delle differenze strutturali nelle due lingue. Infine, la categoria F/Sty/Idiom indica le problematiche relative alla traduzione di espressioni idiomatiche. Per quest'ultima categoria, di seguito vengono riportati due esempi, uno dalla TA prodotta dal modello *fine-tuned* e l'altro da DeepL.

Esempio n. 34 (ProgFT_Testo1 – F/Sty/Idiom)

TO: Dieses Jahr bildet damit einen Einschnitt in der Art, wie [...]

TAFT: Quest'anno forma così un taglio nel modo in cui [...]

PEFT: Quest'anno segna così una svolta nel modo in cui [...]

Esempio n. 35 (Prog1_Testo6 – F/Sty/Idiom)

TO: Dieses Jahr bildet damit einen Einschnitt in der Art, wie [...]

TA: Quest'anno segna quindi una svolta nel modo in cui [...]

Negli esempi riportati, il modello *fine-tuned* rende l'espressione tedesca in modo letterale ("forma un taglio"), senza rispettare il valore idiomatico del segmento, mentre DeepL propone una resa più aderente al significato originale ("segna una svolta"), preservando la funzione semantica e pragmatica dell'espressione nel testo scientifico.

Si rileva lo stesso numero di occorrenze nei due testi tradotti automaticamente per le categorie A/SouEr e A/MisTra/Gend, con un'occorrenza ciascuna. Quest'ultima categoria indica gli errori relativi alla mancata applicazione del linguaggio inclusivo, che nella scrittura scientifica è ormai uno standard consolidato. La categoria A/SouEr, invece, riguarda gli errori già presenti nel testo di partenza e in questo caso consiste nella collocazione errata del numero della nota a piè di pagina (cfr. esempio n. 36). Si tratta di un problema tecnico legato al file traducibile generato da Trados Studio: il numero inserito nel testo come rimando alla nota a piè di pagina viene posizionato in basso invece che in alto come apice. Tale errore viene trasferito così nel testo tradotto, sia dal modello *fine-tuned* sia da DeepL, pur non essendo presente nel file originale.

Esempio n. 36 (ProgFT_Testo1 – A/SouEr)

TO: [...] und die ergiebigsten Ergebnisse versprechen.1

TAFT: [...] e promettono i risultati più significativi.1

PEFT: [...] e promettono i risultati più significativi.¹

Complessivamente, la differenza più rilevante tra i due testi tradotti automaticamente riguarda la categoria F/Sty/Disfl, che segnala le problematiche di scorrevolezza della lingua di arrivo e per la quale il sistema *fine-tuned* registra un numero considerevolmente più elevato di occorrenze. Tuttavia, questo modello sembra dare risultati migliori in riferimento ad alcuni aspetti tipici del linguaggio scientifico, come la realizzazione delle nominalizzazioni (F/SLP/Nomin), la strutturazione dei riferimenti bibliografici (F/Coh/Cit) e l'uso delle forme impersonali (F/SLP/ForImp). Si riscontra anche una gestione più accurata della terminologia, in particolare per quei termini che non richiedono una traduzione (A/Termi/DNT), e un minor ricorso all'inglese come lingua *pivot* (A/ForLa). Anche gli errori di scelta lessicale non adeguata al genere e all'ambito specialistico (F/Lex/LexCho) risultano più contenuti rispetto a DeepL, pur rappresentando ancora una criticità. Anche il modello *fine-tuned* presenta tuttavia anche diverse categorie problematiche: in alcuni casi il numero di errori è uguale a quello di DeepL, ad esempio G/Orth/Punct (errori di punteggiatura), F/De/Layo (errori di strutturazione del testo) e A/MisExpl (contenuto non adeguato al lettore del testo di arrivo), mentre in altri risulta superiore (ad esempio G/Orth/Capit per gli errori di uso della lettera maiuscola e A/Mech/CharFor per gli errori di trasferimento della formattazione). Nonostante ciò, i risultati presentati appaiono incoraggianti e costituiscono una solida base per ulteriori progressi volti a garantire una traduzione di migliore qualità del linguaggio scientifico tedesco in italiano nel contesto linguistico.

8.1.2 Analisi dell'*effort* temporale e tecnico

Il *post-editing* per il testo generato dal sistema *fine-tuned* è stato articolato in sei sessioni, con durate individuali di 01:11:06, 04:17:00, 02:53:35, 00:45:29, 04:10:42 e 02:15:46⁸⁰, per un totale complessivo di 15:33:38⁸¹. Il *post-editing* del testo tradotto da DeepL ha richiesto invece un totale complessivo di 13:34:09.

L'esame congiunto delle durate del *post-editing* e dell'estensione del testo mostra come la lunghezza del materiale tradotto non costituisca un indicatore sufficiente per stimare l'*effort* necessario alla correzione. Pur riferendosi in entrambi i casi allo stesso volume testuale (8.454 *tokens*), i tempi complessivi destinati al *post-editing* divergono in modo significativo. Tale scarto temporale suggerisce che non sia tanto la quantità di testo a influire sul carico di correzione, quanto piuttosto la qualità dell'output prodotto.

Il confronto tra il numero di errori annotati e il tempo dedicato al *post-editing* conferma questa tendenza. Il testo derivante dal modello *fine-tuned* presenta un totale di 1.048 errori, a fronte dei 800 individuati nel testo ottenuto tramite DeepL. La maggiore densità di interventi necessari nel primo caso risulta coerente con la durata più elevata delle sessioni di *post-editing*. Tuttavia, non è la sola quantità degli errori a determinare l'*effort* complessivo: la loro tipologia, gravità e distribuzione interna rappresentano fattori decisivi.

Sulla base del rapporto tra numero di *tokens* e tempo impiegato per le correzioni, è possibile ricavare una stima della produttività del *post-editor*: nel caso del modello *fine-tuned*, essa si attesta a circa 540 *tokens*/ora, mentre per il testo prodotto da DeepL raggiunge un valore approssimativo di 620 *tokens*/ora. Questa differenza evidenzia che l'output di DeepL risulta globalmente più agevole da correggere, richiedendo un numero inferiore di interventi e, presumibilmente, una revisione meno impegnativa. Il modello *fine-tuned*, al contrario, mostra una qualità più instabile, che impone un intervento umano più sistematico e approfondito.

Nel complesso, i risultati confermano che la qualità dell'output della TA è il principale determinante del *post-editing effort*. L'estensione del testo o il semplice conteggio degli errori rappresentano parametri indicativi ma non esaustivi, mentre la natura delle problematiche traduttive e la qualità di partenza emergono come i fattori che più influenzano i tempi di correzione.

Dall'analisi del report di Post-Edit Compare, emerge che il testo prodotto dal modello *fine-tuned* comprende 370 segmenti per un totale di 6.970 parole. Di questi, 353 segmenti provengono direttamente dall'output del sistema di TA, mentre 17 sono stati aggiunti manualmente in assenza di una proposta automatica. Per quanto riguarda la traduzione generata da DeepL, il numero complessivo di segmenti è leggermente superiore (375), pur a fronte dello stesso totale di parole

⁸⁰ Tutte le durate riportate in questo capitolo sono espresse nel formato ore:minuti:secondi (hh:mm:ss).

⁸¹ Tutte le sessioni di *post-editing* riportate nel presente capitolo sono state intervallate ogni 01:30:00/01:50:00 con delle pause di durata variabile, non conteggiate nel tempo totale di lavoro effettivo.

(6.970). Tale discrepanza è verosimilmente attribuibile a differenze nei criteri di segmentazione adottati dai due sistemi. Anche in questo caso, la quasi totalità del testo (364 segmenti) deriva dalla traduzione automatica, mentre 11 segmenti sono stati prodotti manualmente. In nessuno dei due testi si osserva il contributo della memoria di traduzione, circostanza che conferma come il materiale esaminato corrisponda integralmente all’output originario dei sistemi di TA. Tale risultato è riconducibile al funzionamento di Post-Edit Compare, che opera su una versione iniziale unica del progetto, salvata prima dell’avvio del *post-editing*, quando la memoria era ancora priva di contenuti. Lo strumento, pertanto, analizza esclusivamente la versione di partenza e non ricalcola eventuali *matches* generati in seguito.

Nella tabella seguente sono riportati i valori di PEMp (*Post-Edit Modifications percent*), suddivisi in sei fasce. Per ciascuna fascia, le colonne dedicate a segmenti, parole e percentuali riportano i dati relativi al modello *fine-tuned* e a DeepL, mettendo in evidenza differenze significative nella qualità dell’output, misurata sulla base dell’entità delle modifiche richieste in fase di *post-editing*. Tale tabella offre una rappresentazione comparativa chiara e immediata della qualità dei due sistemi e fornisce al contempo una panoramica dettagliata della distribuzione degli interventi.

Fasce di PEMp	Segmenti		Parole		Percentuali	
Sistema di TA	Modello <i>fine-tuned</i>	DeepL	Modello <i>fine-tuned</i>	DeepL	Modello <i>fine-tuned</i>	DeepL
100%	15	75	310	1.164	4,67%	16,70%
95-99%	42	47	814	933	12,27%	13,39 %
85-94%	88	61	1951	1.346	29,41%	19,31%
75-84%	71	51	1323	1.158	19,94%	16,61%
50-74%	60	45	1139	813	17,17%	11,94%
New (0-49%)	70	96	1097	1.556	16,54%	22,32%

Tabella 12: Distribuzione comparativa delle fasce di PEMp per il testo del modello *fine-tuned* n. 1

Nel complesso, entrambe le traduzioni presentano una qualità non omogenea, ma con andamenti distinti che delineano due profili di comportamento diversi.

Nella fascia 100%, che raccoglie i segmenti riutilizzati integralmente, la distanza tra i due sistemi è marcata: il modello *fine-tuned* presenta soltanto il 4,67% delle parole (15 segmenti), mentre DeepL raggiunge 75 segmenti, pari al 16,70% delle parole. Questa differenza segnala che DeepL produce con maggiore frequenza output già utilizzabili senza necessità di intervento.

Nella fascia immediatamente successiva (95-99%), relativa alle modifiche minime, DeepL mantiene un leggero vantaggio: 47 segmenti e il 13,39% delle parole contro i 42 segmenti e il

12,27% del modello *fine-tuned*. Pertanto, in questa fascia, entrambi i sistemi offrono soluzioni sostanzialmente affidabili, che richiedono solo interventi di correzione molto limitati.

La fascia 85-94% rappresenta la quota più ampia per il modello *fine-tuned* con il 29,41% delle parole (88 segmenti). Ciò indica che poco meno di un terzo del suo output ha richiesto correzioni di lieve entità, situandosi spesso in una zona intermedia che implica comunque un intervento umano non trascurabile. Per DeepL, invece, la fascia 85-94% è la seconda più ampia con circa un quinto del suo output (19,31% delle parole, 61 segmenti).

Le fasce 75-84% e 50-74% confermano questa tendenza: il modello *fine-tuned* mostra valori più elevati (19,94% e 17,17%), mentre DeepL si attesta rispettivamente al 16,61% e all'11,94%. In questi casi, il sistema adattato richiede più spesso interventi di entità medio-alta, a testimonianza di una qualità più irregolare rispetto all'output di DeepL.

La fascia *New* (0-49%) presenta un andamento opposto rispetto alle due fasce precedenti. DeepL registra la sua percentuale più elevata con il 22,32% delle parole (96 segmenti), mentre il modello *fine-tuned* raggiunge il 16,54% (70 segmenti). Questo dato mostra che DeepL, pur distinguendosi nelle fasce qualitativamente migliori, tende a fallire in modo più netto quando incorre in difficoltà, producendo segmenti che richiedono una riscrittura quasi totale o integrale.

Nel complesso, i dati restituiscono un quadro articolato: DeepL si distingue nelle fasce di qualità più elevata, poiché fornisce una quantità maggiore di segmenti pienamente riutilizzabili o che necessitano di interventi minimi. Tuttavia, questa maggiore capacità di produrre traduzioni di qualità pubblicabile o quasi è accompagnata da una presenza più marcata di segmenti nella fascia *New* (0-49%), per i quali le proposte della TA si rivelano del tutto inadeguate o quasi. Il modello *fine-tuned*, invece, concentra una parte significativa del proprio output nella fascia 85-94%. Tale comportamento si traduce in una prevalenza di segmenti che richiedono interventi di media portata, a fronte di un numero più ridotto di casi estremi rispetto a DeepL. Questi risultati delineano due profili qualitativi distinti: DeepL polarizza il proprio output tra segmenti molto validi e segmenti estremamente problematici, mentre il modello *fine-tuned* si colloca più stabilmente in una zona intermedia che richiede comunque interventi di correzione.

8.2 Testo per il progetto n. 2 (8.484 tokens)

Corrispondente al testo n. 7 del progetto trans-kom

Agnetta, Marco. 2021. «Kompensation und Redundanz. Zwei Prinzipien der Text-Bild-Beziehung in leicht verständlichen Kommunikaten». *trans-kom* 14 (2): 257–77.

8.2.1 Analisi degli errori

Categoria di errore	Numero di annotazioni nel modello fine-tuned	Numero di annotazioni in DeepL
F/Coh/InPu	210	134
F/Lex/LexCho	167	98
A/MisTra/SemUn	162	140
F/Sty/Disfl	147	74
A/Mech/Punct	58	30
A/Termi/SemUn	43	51
A/Omis	40	24
A/Mech/CharFor	40	2
A/MisTra/WSD	40	19
A/DNT	19	17
G/Orth/Punct	19	20
A/MisOrd/WO	17	11
A/Oth	16	6
A/Untra	14	15
A/Termi/DNT	14	7
F/Sty/Reg	12	9
A/MisTra/POS	12	8
G/Orth/Spel	12	1
A/ForLa	11	14
A/MisTra/Part	11	14
G/WF/Agree	11	5
A/MisExpl	11	9
G/Orth/WhiSpa	8	10
A/MisOrd/MW	8	7
A/MisTra/WroPr	8	2
F/Lex/Fue	7	3
A/SouEr	5	7
A/Add	4	9
F/Sty/Idiom	3	2
G/Orth/Capit	2	16
F/SLP/ForImp	2	7
F/Coh/CoRef	3	4
A/MisTra/Gend	2	2
F/Coh/Cit	1	3
A/MWE/Komp	1	-
G/WF/VF	1	1
F/SLP/Nomin	-	12
A/Mech/Capit	-	1

Tabella 13: Analisi quantitativa comparativa degli errori per il testo del modello *fine-tuned* n. 2

Nel testo tradotto mediante il sistema *fine-tuned* sono stati rilevati 552 errori di *Fluency*, 536 errori di *Accuracy* e 53 errori di *Grammar*, per un totale di 1.141 annotazioni. Il numero di occorrenze per la macrocategoria *Fluency* supera di poco quello della macrocategoria *Accuracy*. La

categoria che registra più risultati è F/Coh/InPu, con 210 occorrenze, che segnala i casi di uso incoerente di segni ortografici, quali l’apostrofo e le virgolette (cfr. esempio n. 37).

Esempio n. 37 (ProgFT_Testo2 – F/Coh/InPu)

TAFT: [...] ridondanza "sinsemantica" [...]

PEFT: [...] ridondanza "sinsemantica" [...]

Nel testo tradotto con il sistema di TA generico DeepL, invece, si riscontrano 346 errori di *Fluency*, 395 di *Accuracy* e 53 di *Grammar*. Il totale degli errori annotati ammonta a 794. In questo caso, la quota più consistente di annotazioni rientra nella macrocategoria *Accuracy*. La categoria A/MisTra/SemUn, riguardante la traduzione semanticamente errata, risulta la maggiormente rappresentata, mentre la categoria F/Coh/InPu ricopre il secondo posto. Nel complesso il sistema *fine-tuned* presenta un numero più elevato di errori sia a livello delle singole macrocategorie che nel conteggio totale, con l’unica eccezione della macrocategoria *Grammar*, per la quale i due sistemi registrano lo stesso numero di errori.

Proseguendo per numero di occorrenze nel testo generato dal modello *fine-tuned*, si registrano 167 errori per la categoria F/Lex/LexCho, che, come si è già specificato, è particolarmente significativa per la traduzione del linguaggio scientifico, in quanto riguarda l’impiego di lessemi non idonei al genere testuale in esame. Nel testo generato da DeepL si registrano 98 occorrenze. Questi dati evidenziano pertanto una maggiore frequenza nel modello *fine-tuned*. Di seguito viene presentato un esempio tratto dal testo generato dal sistema adattato.

Esempio n. 38 (ProgFT_Testo2 – F/Lex/LexCho)

TO: Die Summe der zielsprachlichen Mittel ist [...]

TAFT: La somma delle risorse linguistiche target è [...]

PEFT: La somma delle risorse della lingua di arrivo è [...]

Nell’esempio riportato, l’errore riguarda l’aggettivo “zielsprachlichen” tradotto come “linguistiche target”, presentando quindi l’anglicismo “target”. In fase di *post-editing*, è stata impiegata l’espressione “della lingua di arrivo”, formulazione maggiormente conforme al linguaggio scientifico impiegato in ambito traduttologico e linguistico.

Nel testo del modello *fine-tuned* si riscontrano anche le seguenti categorie di errori: A/MisTra/SemUn con 161 occorrenze (traduzione semanticamente errata) e F/Sty/Disfl con 147 (problematiche di scorrevolezza della lingua di arrivo). Queste categorie registrano i maggiori risultati anche nel testo tradotto con DeepL: 163 errori per A/MisTra/SemUn e 74 per F/Sty/Disfl. I

valori risultano dunque più elevati nel testo generato dal modello *fine-tuned*, dove la categoria F/Sty/Disfl risulta particolarmente critica (cfr. esempio n. 39).

Esempio n. 39 (ProgFT_Testo2 – F/Sty/Disfl)

TO: [...] die sie aber nach den angesprochenen Lernprozessen nicht mehr benötigen.

TAFT: [...] ma di cui non hanno più bisogno dopo i processi di apprendimento di cui si è parlato.

PEFT: [...] ma di cui non hanno più bisogno dopo i processi di apprendimento menzionati.

La proposta formulata dal modello adattato presenta una resa ridondante di “di cui” e poco scorrevole. In fase di *post-editing*, la ripetizione è stata eliminata e la struttura resa più lineare, risultando maggiormente idonea al linguaggio scientifico. Per gli errori di A/MisTra/SemUn, che indicano la mancata corrispondenza semantica tra testo di partenza e testo di arrivo, viene riportato di seguito un esempio tratto dal testo del modello *fine-tuned*.

Esempio n. 40 (ProgFT_Testo2 – A/MisTra/SemUn)

TO: Die nachfolgenden Überlegungen schließen demnach die genuine Produktion [...] aus.

TA: Le riflessioni che seguono, quindi, escludono [...] la produzione genovese [...].

PE: Le riflessioni che seguono, quindi, escludono [...] la produzione vera e propria [...].

Nell'esempio n. 40 il sistema *fine-tuned* interpreta erroneamente “genuine” come “genovese”, producendo una resa semantica errata che compromette la trasmissione del significato del testo di partenza. L'intervento di *post-editing* ripristina il significato originale con “vera e propria”.

Un'ulteriore categoria di errori con un numero significativo di occorrenze è A/Mech/Punct, che segnala casi di errato trasferimento della punteggiatura e registra 58 errori nel testo tradotto dal modello *fine-tuned* e 30 in quello reso da DeepL.

Esempio n. 41 (ProgFT_Testo2 – A/Mech/Punct)

TO: [...] als ‘dieselbe Sprache’ erachtet wird [...]

TAFT: [...] viene considerato come la stessa lingua! [...]

PEFT: [...] viene considerato come ‘la stessa lingua’ [...]

Esempio n. 42 (Prog2_Testo7 – A/Mech/Punct)

TO: ‘Synsemantische Redundanz’ ist [...]

TA: La ridondanza sinsemantica è [...]

PE: La ‘ridondanza sinsemantica’ è [...]

Negli esempi riportati, gli errori di A/Mech/Punct riguardano il mancato trasferimento delle virgolette. Nel primo esempio, oltre all'assenza della virgoletta di apertura, si riscontra anche un errore di tipo F/Coh/InPu, dovuto all'uso incoerente delle virgolette: il modello utilizza la virgoletta dritta in luogo della virgoletta tipografica. Nel secondo esempio, invece, le virgolette sono del tutto assenti. Questi errori compromettono la fedeltà al testo di partenza e la corretta resa tipografica nella lingua di arrivo, come evidenziato dalle correzioni apportate in fase di *post-editing*.

L'analisi dei testi prodotti dai due sistemi mostra una presenza significativa di errori terminologici, che si distribuiscono in modo diverso tra le categorie considerate. Nel testo generato dal sistema *fine-tuned* sono stati individuati 43 errori di A/Termi/SemUn, relativi alla resa errata di termini, mentre nel testo tradotto da DeepL ne sono stati individuati 51. Per quanto riguarda la categoria A/Termi/DNT, che segnala termini che non devono essere tradotti, si riscontrano 16 errori nel testo del modello *fine-tuned* e 10 in quello di DeepL. Gli esempi seguenti illustrano alcuni casi significativi di errori terminologici nei due sistemi.

Esempio n. 43 (ProgFT_Testo2 – A/Termi/SemUn)

TO: Die Abfassung von Texten in Leichter Sprache [...]

TAFT: La stesura dei testi in lingua leggera [...]

PEFT: La stesura dei testi in linguaggio facile [...]

Esempio n. 44 (Prog2_Testo7 – A/Termi/SemUn)

TO: Die Abfassung von Texten in Leichter Sprache [...]

TA: La scrittura di testi in linguaggio semplice [...]

PE: La stesura dei testi in linguaggio facile [...]

Gli esempi n. 43 e n. 44 presentano un errore di A/Termi/SemUn riguardante lo stesso segmento tradotto nel primo esempio dal modello *fine-tuned* e nel secondo esempio da DeepL. Entrambi i sistemi traducono “Leichte Sprache” con soluzioni terminologiche non adeguate (“lingua leggera”, “linguaggio semplice”), che non restituiscono il valore specialistico del termine. In particolare, nella proposta di traduzione il modello *fine-tuned* utilizza “lingua” invece di “linguaggio”, introducendo un'ulteriore discrepanza rispetto alla convenzione terminologica italiana. Come evidenziato da Crestani (2022: 243) e come è stato ampiamente spiegato nell'esempio n. 48 del capitolo 6, la lingua italiana presenta una notevole variabilità denominativa per questo concetto, che rende necessaria una scelta traduttiva consapevole e coerente. In accordo con Crestani (2022) che propone il termine “linguaggio facile italiano” per riferirsi alla variante italiana della *leichte Sprache*, in tale contributo si adotta il termine italiano “linguaggio facile” per indicare la varietà indicata nel suo

complesso, indipendentemente dalla lingua standard di riferimento. Le rese proposte dai due sistemi risultano quindi inappropriate, poiché non riconoscono la specificità del termine tedesco né la convenzione terminologica della lingua italiana.

Esempio n. 45 (ProgFT_Testo2 – A/Termi/DNT)

TO: Leichte Sprache gehört dem System der deutschen Sprache zu [...]

TAFT: La lingua leggera appartiene al sistema della lingua tedesca [...]

PEFT: La Leichte Sprache appartiene al sistema della lingua tedesca [...]

Esempio n. 46 (Prog2_Testo7 – A/Termi/DNT)

TO: Leichte Sprache gehört dem System der deutschen Sprache zu [...]

TA: La lingua facile appartiene al sistema della lingua tedesca [...]

PE: La Leichte Sprache appartiene al sistema della lingua tedesca [...]

Anche gli esempi n. 45 e n. 46 si riferiscono allo stesso segmento del testo di partenza e presentano il medesimo errore di A/Termi/DNT: nel primo caso la traduzione è prodotta dal modello *fine-tuned*, nel secondo da DeepL. Entrambi i sistemi traducono il termine “Leichte Sprache” in italiano, nonostante in questo contesto la traduzione non sia necessaria. In questo caso “Leichte Sprache” designa infatti una varietà linguistica del tedesco sviluppata per rendere i testi scritti accessibili a persone con difficoltà di apprendimento, sulla base di regole e linee guida codificate (Crestani 2024: 200-201). Il termine indica quindi in tal caso una denominazione istituzionalizzata e culturalmente specifica, che non trova un equivalente univoco e stabilizzato nella lingua italiana. Per queste ragioni, in fase di *post-editing* si è scelto di mantenere la forma tedesca originale: ciò consente di preservare la precisione concettuale del termine, di evitare ambiguità derivanti dalla variabilità delle denominazioni italiane e di mantenere il riferimento esplicito al quadro normativo e sociolinguistico entro cui il termine è definito.

I dati riportati nella tabella evidenziano altre categorie di *Accuracy* che registrano un numero significativo di occorrenze nel testo generato dal modello *fine-tuned*: A/Omis, A/Mech/CharFor, A/MisTra/WSD e A/DNT. Tra queste, la categoria A/Omis, che segnala i segmenti omessi nel testo di arrivo, risulta particolarmente rilevante, con 40 occorrenze nel testo del modello *fine-tuned* e 24 nel testo di DeepL. Nel testo del sistema adattato, le omissioni interessano prevalentemente singole parole, la cui assenza può compromettere il senso della frase, come mostra l’esempio n. 47.

Esempio n. 47 (ProgFT_Testo2 – A/Omis)

TO: [...] sofern sie sich auf ein bereits existierendes Kommunikat [...] zurückführen lassen, im strengen Sinne also intertextuell bzw. intermedial sind.

TAFT: a condizione che possano essere ricondotte a una comunicazione già esistente [...] e quindi, in senso stretto, intertestuali o intermediali.

PEFT: a condizione che possano essere ricondotte a una comunicazione già esistente [...] e quindi, in senso stretto, siano intertestuali o intermediali.

Nell'esempio n. 47 è stata omessa la copula “siano”, fondamentale per completare la frase in italiano. L'assenza di questa parola singola rende la frase incompleta e compromette la comprensibilità del segmento rispetto al testo di partenza. Nel testo tradotto da DeepL, invece, le omissioni risultano più variabili e possono coinvolgere singole parole (cfr. esempio n. 48), gruppi di parole o addirittura segmenti interi, con un impatto più marcato sulla completezza informativa del testo.

Esempio n. 48 (Prog2_Testo7 – A/Omis)

TO: Zuweilen wird in der zeitgenössischen Semiotik der Code- bzw. Kodebegriff gemieden, [...]

TA: Nella semiotica contemporanea, il concetto di codice viene talvolta evitato [...]

PE: Nella semiotica contemporanea, il concetto di code o codice viene talvolta evitato [...]

Nell'esempio n. 48 l'omissione riguarda la formulazione “Code- bzw.”, che viene ripristinata correttamente nella proposta di *post-editing*. L'assenza di uno o più lessemi può ridurre la precisione informativa della traduzione e alterare il significato complessivo del testo, confermando la classificazione come errore di *Accuracy*.

La categoria A/Mech/CharFor riguarda gli errori nel trasferimento della formattazione (cfr. esempio n. 8 in 8.1.1). Tale categoria si presenta particolarmente problematica nel testo generato dal modello *fine-tuned*, che presenta 40 occorrenze, a fronte delle sole 2 individuate nel testo prodotto da DeepL. La categoria A/DNT segnala invece i casi in cui elementi che non richiederebbero traduzione vengono erroneamente tradotti. In questo caso, si riscontrano 19 errori nel testo del modello *fine-tuned* e 17 in quello di DeepL. Di seguito sono riportati due esempi: l'esempio n. 49 tratto dal testo del sistema adattato e l'esempio n. 50 per l'output di DeepL.

Esempio n. 49 (ProgFT_Testo2 – A/DNT)

TO: [...] in ihrer grundlegenden Studie “Übersetzung und Kompensation” [...]

TAFT: [...] nel suo studio fondamentale “traduzione e compensazione” [...]

PEFT: [...] nel suo studio fondamentale “Übersetzung und Kompensation” [...]

Esempio n. 50 (Prog2_Testo7 – A/DNT)

TO: [...] in seinem vielzitierten Aufsatz “On linguistic aspects of translation” [...]

TA: [...] nel suo saggio pluricitato "Sugli aspetti linguistici della traduzione" [...]

PE: [...] nel suo saggio pluricitato "On linguistic aspects of translation" [...]

Negli esempi n. 49 e n. 50 si riscontra un errore di tipo A/DNT, riguardante in particolare titoli di opere, nel primo esempio in tedesco e nel secondo in inglese, che sarebbero dovuti rimanere invariati. La traduzione automatica in italiano altera la fedeltà al testo di partenza e i riferimenti bibliografici, mentre l'intervento di *post-editing* preserva la forma originale dei titoli, garantendo accuratezza e aderenza al contenuto originario.

Gli errori della categoria A/MisTra/WSD, infine, riguardano scelte semantiche inadeguate riferite a lessemi polisemici, per i quali il sistema di TA seleziona un significato non pertinente al contesto. Questa categoria registra 40 occorrenze nel testo del modello *fine-tuned* e 19 nel testo tradotto con DeepL. L'esempio seguente ne presenta un caso tratto dal testo del sistema adattato.

Esempio n. 51 (ProgFT_Testo2 – A/MisTra/WSD)

TO: [...] seine Unterscheidung von Übersetzung und Bearbeitung [...]

TAFT: [...] la sua distinzione tra traduzione e lavorazione [...]

PEFT: [...] la sua distinzione tra traduzione e rielaborazione [...]

Nell'esempio n. 51, il lessema "Bearbeitung", polisemico in tedesco, viene reso impropriamente con "lavorazione". In fase di *post-editing*, la scelta lessicale è stata corretta con "rielaborazione", soluzione pertinente al contesto e al significato specialistico.

Per quanto riguarda la macrocategoria *Grammar*, i due testi tradotti automaticamente presentano lo stesso numero complessivo di errori, ma con una distribuzione significativamente diversa tra le varie sottocategorie. Nel testo prodotto dal modello *fine-tuned* si rilevano i seguenti dati: G/Orth/Punct (errori di punteggiatura) con 19 occorrenze, G/Orth/Spel (errori ortografici) con 12 occorrenze, G/WF/Agree (errori di accordo tra parole) 11 occorrenze, G/Orth/WhiSpa (errori di spaziatura) con 8 occorrenze, G/Orth/Capit (errori di uso della lettera maiuscola) con 2 occorrenze e G/WF/VF (errore di tempo e forma verbale) con un'occorrenza. Nel testo generato da DeepL, invece, gli errori si distribuiscono nel seguente modo: G/Orth/Punct con 20 errori, G/Orth/Capit con 16, G/Orth/WhiSpa con 10, G/WF/Agree con 5 e G/Orth/Spel e G/WF/VF con un errore ciascuna. Nel complesso, dunque, entrambi i sistemi condividono le stesse problematiche, ma ciascuno tende a concentrare gli errori in modo diverso: il modello *fine-tuned* produce un numero più elevato di errori di concordanza (G/WF/Agree, cfr. esempio n. 52) e ortografia (G/Orth/Spel, cfr. esempio n. 53); mentre DeepL presenta una maggiore incidenza di errori nell'uso della maiuscola

(G/Orth/Capit, cfr. esempio 54), della punteggiatura (G/Orth/Punct) e della spaziatura (G/Orth/WhiSpa).

Esempio n. 52 (ProgFT_Testo2 – G/WF/Agree)

TO: [...] über die beiden gleichnamigen Gedichte von Goethe [...]

TAFT: [...] sulle due poesie omonima di Goethe [...]

PEFT: [...] sulle due poesie omonime di Goethe [...]

L'esempio n. 52 mostra un tipico errore di concordanza (G/WF/Agree): l'aggettivo "omonima" non concorda con il sostantivo plurale "poesie" a cui fa riferimento. L'intervento di *post-editing* ripristina la corretta forma "omonime", garantendo la coerenza morfosintattica.

Esempio n. 53 (ProgFT_Testo2 – G/Orth/Spel)

TO: [...] für entfernte diatopische und diachrone Varietäten [...]

TAFT: [...] per varietà diatopice e diacroniche distanti [...] sulle due poesie omonima di Goethe

PEFT: [...] per varietà diatopiche e diacroniche distanti [...]

L'esempio n. 53 presenta un errore di G/Orth/Spel: nella resa del sintagma tedesco "diatopische und diachrone Varietäten", il modello *fine-tuned* produce la forma errata "diatopice" invece della forma corretta "diatopiche", generando un errore ortografico che compromette la correttezza formale del testo.

Esempio n. 54 (Prog2_Testo7 – G/Orth/Capit)

TO: [...] soll folgendermaßen vorgegangen werden: In einem ersten Abschnitt [...]

TA: [...] si procederà come segue: Nella prima sezione [...]

PE: [...] si procederà come segue: nella prima sezione [...]

Nell'esempio summenzionato, il sistema generico DeepL riproduce la convenzione ortografica tedesca, secondo la quale dopo i due punti si usa la lettera maiuscola quando segue una frase completa. Il sistema di TA non adatta quindi questa struttura alla norma italiana, che non richiede invece l'uso della minuscola.

Proseguendo nell'analisi, si individuano le seguenti categorie di errore: A/MisOrd/WO per gli errori riconducibili alle variazioni dell'ordine degli elementi all'interno della frase che alterano il significato e compromettono la trasmissione accurata del messaggio del testo di partenza (17 errori nel testo del modello *fine-tuned* e 11 in quello di DeepL); A/Untra riguardante i segmenti non tradotti (14 occorrenze nel testo tradotto dal modello *fine-tuned* e 15 nel testo generato da DeepL); infine, la categoria A/Oth per gli errori non riconducibili ad altre categorie di errori (15 occorrenze

nel testo generato dal modello *fine-tuned* e 6 nel testo prodotto da DeepL). Gli esempi seguenti illustrano alcuni errori ricorrenti per quest'ultima categoria.

Esempio n. 55 (ProgFT_Testo2 – A/Oth)

TO: [...] (Maaß 2015: 12)

TAFT: [...] (Mass 2015: 12)

PEFT: [...] (Maaß 2015: 12)

L'esempio n. 55 riguarda la trasmissione errata di un nome proprio. Si tratta di una criticità che emerge con una certa frequenza nel testo del modello *fine-tuned*, che comporta una modifica del riferimento bibliografico e può interferire con la tracciabilità delle fonti.

Esempio n. 56 (ProgFT_Testo2 – A/Oth)

TO: Verwenden Sie immer Bilder, die klar und leicht verständlich sind. Die Bilder müssen eindeutig zu dem Text passen.

TAFT: Usa sempre le immagini, che sono chiare e facilmente comprensibili. Le immagini devono essere chiaramente in linea con il testo.

PEFT: Scegliete sempre illustrazioni chiare, facili da capire e adatte al brano che cercano di spiegare.

Esempio n. 57 (Prog2_Testo7 – A/Oth)

TO: Verwenden Sie immer Bilder, die klar und leicht verständlich sind. Die Bilder müssen eindeutig zu dem Text passen.

TA: Utilizzate sempre le immagini, che siano chiari e di facile comprensione. Le immagini devono corrispondere chiaramente al testo.

PE: Scegliete sempre illustrazioni chiare, facili da capire e adatte al brano che cercano di spiegare.

Gli esempi n. 56 e 57 mostrano lo stesso tipo di errore, presente in entrambi i testi tradotti automaticamente. La categoria A/Oth è stata impiegata in questo caso per segnalare la traduzione errata di una citazione: il testo tedesco citato è tratto dalle linee guida europee per la produzione di contenuti facili da leggere e da comprendere⁸², per le quali esiste una versione ufficiale in lingua italiana. Né il modello *fine-tuned* né DeepL hanno riconosciuto la natura del segmento e non hanno impiegato la corrispondente versione italiana ufficiale. In fase di *post-editing*, pertanto, la traduzione letterale prodotta dai sistemi è stata sostituita con il testo corretto riprodotto fedelmente

⁸² Per consultare la versione tedesca delle linee guida, si veda: https://www.inclusion-europe.eu/wp-content/uploads/2017/06/DE_Information_for_all.pdf (ultima consultazione: 20/12/2025); per la versione italiana si veda invece: https://www.inclusion-europe.eu/wp-content/uploads/2017/06/IT_Information_for_all.pdf (ultima consultazione: 20/12/2025).

dalle linee guida italiane. Inoltre, negli esempi, la traduzione prodotta dai sistemi presenta alcune problematiche ricorrenti: la scelta della persona verbale per il modello *fine-tuned* (“Usa” in sostituzione “Scegliete”) non corrisponde allo stile previsto; la ripetizione di “Le immagini” da parte di entrambi i sistemi rende il testo ridondante; l’aggettivo “chiari” per DeepL mostra un errore di concordanza con il sostantivo. Tali aspetti, insieme al mancato riconoscimento della citazione ufficiale, compromettono la resa adeguata del segmento.

Per quanto attiene gli elementi caratterizzanti del linguaggio scientifico, si segnala in primo luogo la categoria F/Sty/Reg che indica le problematiche di registro inadeguato. Tale categoria presenta 12 errori nel testo generato dal sistema *fine-tuned* e 9 nel testo prodotto da DeepL. Pertanto, il rispetto di tale requisito si rivela più problematico nel sistema adattato, come mostra il seguente esempio 58.

Esempio n. 58 (ProgFT_Testo2 – F/Sty/Reg)

TO: Hier genügt der Hinweis darauf, dass [...]

TAFT: In questa sede basta indicare che [...]

PEFT: In questa sede è sufficiente indicare che [...]

Nell’esempio n. 58, l’intervento di *post-editing* ha sostituito la formulazione “basta indicare che” con “è sufficiente indicare che”, uniformando così il registro al tono formale richiesto dal genere testuale trattato.

Proseguendo nell’analisi degli elementi caratterizzanti del linguaggio scientifico, si riscontra la categoria F/Lex/Fue che riguarda il lessico del linguaggio scientifico e più nello specifico gli errori nella traduzione delle *Fügungen*. La corretta resa di queste combinazioni idiomatiche risulta più problematica per il modello *fine-tuned*, nella cui traduzione si rilevano 7 occorrenze (cfr. esempio n. 59). Nel testo di DeepL invece si riscontrano 3 occorrenze.

Esempio n. 59 (ProgFT_Testo2 – F/Lex/Fue)

TO: [...] die bei näherem Hinschauen eine Pluralität an Interpretationen hervorrufen können.

TAFT: [...] che, a guardare più da vicino, possono provocare una pluralità di interpretazioni.

PEFT: [...] che, a un esame più attento, possono provocare una pluralità di interpretazioni.

In riferimento ai fenomeni tipici del linguaggio scientifico, si individuano inoltre le due seguenti categorie: F/SLP/ForImp e F/SLP/Nomin. La categoria F/SLP/ForImp segnala i casi in cui il sistema non ricorre alla forma impersonale e la sostituisce con una costruzione personale. Questa

tipologia di errore si presenta con 2 occorrenze nel testo generato dal modello *fine-tuned* e con 7 in quello prodotto da DeepL.

Esempio n. 60 (ProgFT_Testo2 – F/SLP/ForImp)

TO: Zur Beantwortung dieser Fragen soll folgendermaßen vorgegangen werden: [...]

TAFT: Per rispondere a queste domande, si intende procedere come segue: [...]

Esempio n. 61 (Prog2_Testo7 – F/SLP/ForImp)

TO: Zur Beantwortung dieser Fragen soll folgendermaßen vorgegangen werden: [...]

TA: Per rispondere a queste domande, procederemo come segue [...]

PE: Per rispondere a queste domande, si intende procedere come segue:

Nell'esempio n. 61 emerge la difficoltà del sistema generico DeepL nel riconoscere e riprodurre la costruzione impersonale richiesta dal segmento di partenza. Mentre il modello *fine-tuned* propone una resa conforme alle convenzioni del linguaggio scientifico italiano (“si intende procedere”), DeepL opta per una formulazione personale (“procederemo”), modificando il grado di formalità e la neutralità tipica del linguaggio scientifico. In fase di *post-editing* si è intervenuto, quindi, ripristinando la struttura impersonale, più appropriata al registro richiesto.

La categoria F/SLP/Nomin, invece, riguarda la mancata realizzazione di nominalizzazioni, che contribuiscono alla concisione e all'oggettività. Questo tipo di errore risulta particolarmente frequente nel testo prodotto da DeepL, con 12 occorrenze, mentre non è stato individuato alcun caso nel testo del modello *fine-tuned*.

Esempio n. 62 (ProgFT_Testo2 – F/SLP/Nomin)

TO: Redundanz bezweckt [...] das allmähliche Lernen der Zielgruppe.

TAFT: La ridondanza [...] ha come obiettivo l'apprendimento graduale del gruppo target.

Esempio n. 63 (Prog2_Testo7 – F/SLP/Nomin)

TO: Redundanz bezweckt [...] das allmähliche Lernen der Zielgruppe.

TA: La ridondanza mira a insegnare gradualmente al gruppo target [...].

PE: La ridondanza mira all'apprendimento graduale del gruppo target. [...].

Nell'esempio n. 63, DeepL sostituisce il sintagma nominale “das allmähliche Lernen” con una costruzione verbale (“mira a insegnare”), alterando così lo stile e l'astrattezza tipica del linguaggio scientifico. Contestualmente, la resa di “Lernen” con “insegnare” costituisce un errore semantico (A/MisTra/SemUn), poiché il lessema denomina l'apprendimento e non l'insegnamento,

modificando così il significato originario del segmento. Nell'esempio n. 62, invece, il modello *fine-tuned* mantiene la struttura nominale, generando un output più aderente allo stile richiesto dal linguaggio scientifico.

Proseguendo in ordine decrescente di occorrenze nel testo generato dal modello *fine-tuned*, si individua la categoria A/MisTra/POS. Questa riguarda l'uso della categoria lessicale non corrispondente a quella del testo di partenza, compromettendo la corrispondenza informativa tra testo di partenza e testo di arrivo ma non la correttezza grammaticale della lingua di arrivo (cfr. esempio n. 31 in 8.1). Tale categoria registra 12 occorrenze nel testo del sistema adattato e 8 in quello di DeepL, segnalando quindi una gestione più problematica da parte del modello *fine-tuned*. Successivamente si individuano le categorie A/ForLa e A/MisTra/Part, che presentano una tendenza comune: registrano più errori nel testo generato da DeepL. Più nello specifico, la categoria A/ForLa indica i casi in cui è stata impiegata una lingua di arrivo diversa da quella richiesta o inesistente e registra 11 occorrenze nel testo generato dal modello adattato e 14 nel testo prodotto da DeepL. Gli esempi seguenti illustrano alcuni casi individuati nei testi.

Esempio n. 64 (ProgFT_Testo2 – A/ForLa)

TO: [...] die unter die Kategorie [...] subsumierten Schriftarten [...]

TAFT: [...] i font sucuminati nella categoria [...]

PEFT: [...] i font sussunti nella categoria [...]

Nell'esempio n. 64 l'errore riguarda l'uso della parola "sucuminati": si tratta di un errore appartenente alla categoria A/ForLa, in quanto il modello ha tentato di tradurre il lessema tedesco ("subsumierten") senza riuscire a individuare un equivalente corretto nella lingua di arrivo e producendo quindi un lessema inesistente. In fase di *post-editing*, il lessema è stato corretto con la parola "sussunti", restituendo il significato del testo originale.

Esempio n. 65 (ProgFT_Testo2 – A/ForLa)

TO: Die Aufnahme-, Aufmerksamkeits- und Prozessierungskapazität des Lerners [...]

TAFT: La capacità di ricezione, attenzione e processing dell'apprendente [...]

PEFT: La capacità di ricezione, attenzione ed elaborazione dell'apprendente [...]

Esempio n. 66 (Prog2_Testo7 – A/ForLa)

TO: [...] als 'modes', 'Modi' oder 'Modalitäten' bezeichnet [...]

TA: [...] definite come 'modes', 'modi' o 'modalities'.

PE: [...] definite come 'modes', 'modi' o 'modalità'.

Negli esempi n. 65 e n. 66 si osserva un errore riconducibile alla categoria A/ForLa, poiché i sistemi introducono lessemi inglesi (“processing”, “modalities”) in contesti in cui la frase di arrivo richiede un equivalente italiano. In entrambi i casi, il ricorso ingiustificato all’inglese compromette la coerenza linguistica della traduzione. In fase di *post-editing*, tali forme sono state sostituite con i corrispettivi italiani appropriati (“elaborazione”, “modalità”), ripristinando l’accuratezza e l’aderenza al linguaggio scientifico richiesto.

La categoria A/MisTra/Part, invece, riguarda i casi in cui il significato del segmento di partenza viene trasferito solo parzialmente o in modo impreciso. Sono stati individuati 11 errori nel testo del modello *fine-tuned* e 14 in quello tradotto da DeepL. Di seguito vengono riportati due esempi: l’esempio n. 67 per il testo tradotto da DeepL e l’esempio n. 68 per il testo generato dal modello adattato.

Esempio n. 67 (Prog2_Testo7 – A/MisTra/Part)

TO: [...] denn der Empfänger hat es hier nicht mit sich in der Zeit verändernden Inhalten [...] zu tun [...]

TA: [...] perché il destinatario non si trova di fronte a un contenuto che cambia nel tempo [...]

PE: [...] perché il destinatario non si trova di fronte a contenuti che cambiano nel tempo [...]

Esempio n. 68 (ProgFT_Testo2 – A/MisTra/Part)

TO: Das Regelwerk von Inclusion Europe beinhaltet etwa folgenden Passus zum Einsatz von Bildern [...]

TAFT: Le regole di Inclusion Europe comprendono, ad esempio, i seguenti passaggi sull’uso delle immagini [...]

PEFT: Le regole di Inclusion Europe comprendono, ad esempio, il seguente passaggio sull’uso delle immagini [...]

Negli esempi riportati, gli errori della categoria A/MisTra/Part riguardano un trasferimento parziale del significato, dovuto a una modifica del numero grammaticale del sintagma. Nell’esempio n. 67, DeepL impiega la forma singolare (“un contenuto”) in sostituzione del plurale richiesto dal segmento di partenza. Nel secondo esempio, invece, il modello *fine-tuned* compie l’errore inverso, sostituendo il singolare (“der folgende Passus”) con un plurale. In entrambi i casi, la perdita di corrispondenza tra numero grammaticale e riferimento semantico compromette la precisione del trasferimento e rende necessario un intervento di *post-editing* per ristabilire la corretta relazione tra testo di partenza e testo di arrivo. È opportuno precisare che tali occorrenze non sono state classificate come errori di tipo *Grammar*, poiché la forma prodotta nelle TA risulta

grammaticalmente corretta nella lingua di arrivo. L'errore riguarda piuttosto una mancata fedeltà al testo di partenza, e pertanto rientra nella categoria di *Accuracy*, in quanto compromette la precisione del trasferimento di significato.

Successivamente, si individuano le seguenti categorie di *Accuracy*: A/MisExpl, A/MisOrd/MW e A/MisTra/WroPr. La categoria A/MisExpl segnala i casi di inadeguatezza dei contenuti per il destinatario del testo di arrivo e registra 11 occorrenze nel testo tradotto dal modello *fine-tuned* e 9 nel testo di DeepL. L'esempio n. 69 mostra un errore del testo generato dal modello *fine-tuned* per quest'ultima categoria. L'esempio n. 70, invece, illustra il diverso comportamento di DeepL sul medesimo segmento di partenza, riconducibile a una diversa categoria di errore.

Esempio n. 69 (ProgFT_Testo2 – A/MisExpl)

TO: [...] durch die wertenden Begriffe “Schlecht” und “Gut” [...] kodiert.

TAFT: [...] è codificata dai termini valutativi “cattivo” e “buono” [...]

PEFT: [...] è codificata dai termini valutativi “Schlecht” (cattivo) e “Gut” (buono) [...]

Esempio n. 70 (Prog2_Testo7 – A/ForLa)

TO: [...] durch die wertenden Begriffe “Schlecht” und “Gut” [...] kodiert.

TA: [...] è codificata dai termini valutativi “Bad” (cattivo) e “Good” (buono) [...]

PE: [...] è codificata dai termini valutativi “Schlecht” (cattivo) e “Gut” (buono) [...]

Nel segmento di partenza, i lessemi “Schlecht” e “Gut” non svolgono la funzione di semplici aggettivi, ma fanno riferimento diretto a parole presenti in un'immagine a cui rimanda il testo. Nel modello *fine-tuned* (esempio n. 69), la traduzione dei due termini in “cattivo” e “buono” elimina tale riferimento: un lettore italiano non riuscirebbe a riconoscere la corrispondenza con l'immagine, compromettendo la funzione esplicativa del testo. La traduzione risulta pertanto inadeguata rispetto ai bisogni informativi del destinatario. Nel testo generato da DeepL (esempio n. 70), invece, viene mantenuto il riferimento ai due lessemi presenti nell'immagine che vengono però riportati in inglese (“Bad” e “Good”), forme non presenti né nel testo di partenza né nell'immagine. Pertanto, in questo caso si presenta un errore di A/ForLa, dovuto all'uso improprio della lingua inglese. Inoltre, in entrambi gli esempi si individuano degli errori di F/Coh/InPu, riconducibili all'impiego incoerente delle virgolette.

Le categorie A/MisOrd/MW e A/MisTra/WroPr presentano entrambe più occorrenze nel testo tradotto dal modello adattato: la categoria A/MisTra/WroPr segnala i casi di uso errato di preposizioni, sintagmi preposizionali o congiunzioni, e presenta 8 errori nel testo del sistema *fine-tuned* e 2 nel testo di DeepL; invece, la categoria A/MisOrd/MW riguarda gli errori di attribuzione

di elementi all'interno della frase e registra 8 errori nel testo del modello *fine-tuned* e 7 in quello di DeepL. Entrambe le categorie segnalano errori che alterano la relazione semantica tra gli elementi della frase, senza comprometterne però la correttezza grammaticale. Di seguito viene riportato un esempio per la categoria A/MisOrd/MW.

Esempio n. 71 (ProgFT_Testo2 – A/MisOrd/MW)

TO: [...] bei ausgangssprachlichen Textteilen oder gar Ganztexten [...]

TAFT: [...] nel caso di parti testuali di partenza o addirittura interi testi, [...]

PEFT: [...] nel caso di parti testuali o addirittura interi testi in lingua di partenza [...]

Nel caso dell'esempio n. 71, l'errore di A/MisOrd/MW riguarda l'attribuzione scorretta dell'aggettivo "ausgangssprachlichen", che nella proposta del modello *fine-tuned* viene associato soltanto a "parti testuali", mentre nel testo di partenza si riferisce sia a "parti testuali" che "interi testi". L'informazione viene dunque attribuita in modo parziale e non conforme alla struttura semantica originale, compromettendo la corretta interpretazione del segmento. Tale errore non incide dunque sulla correttezza formale della lingua di arrivo, che risulta grammaticalmente corretta. Oltre a ciò, nel segmento tradotto si osserva anche un errore di A/Omis, dovuto all'omissione del lessema "sprachlichen", che comporta la perdita dell'informazione essenziale.

Le categorie A/SouEr e A/Add mostrano una tendenza comune: entrambe presentano un numero più elevato di occorrenze nel testo tradotto da DeepL rispetto al modello *fine-tuned*. Più nello specifico, la categoria A/SouEr fa riferimento a errori già presenti nel testo di partenza e registra 5 occorrenze nel testo del modello *fine-tuned* e 6 nel testo tradotto da DeepL. Gli errori di questa categoria riguardano l'errata collocazione del numero di rimando alla nota a piè di pagina, posizionato in basso anziché in forma di apice, come già illustrato nell'esempio n. 36 del paragrafo 8.1.1. Si tratta di un problema tecnico legato al file traducibile generato da Trados Studio, che quindi viene trasferito nella traduzione, ma che non è presente nel file originale. Tali errori riguardano allo stesso modo il modello *fine-tuned* sia da DeepL. La lieve differenza quantitativa tra i due testi è dovuta a omissioni selettive, da parte del modello adattato, dei rimandi problematici già presenti nel file di partenza. La categoria A/Add, invece, segnala l'aggiunta non necessaria di parole o segmenti. Nel testo di DeepL si registrano 9 casi, mentre in quello generato dal modello adattato solo 4. Inoltre, la natura delle aggiunte differisce nei due sistemi: nel testo di DeepL possono riguardare più parole o interi sintagmi, mentre nel testo del modello *fine-tuned* interessano principalmente singoli lessemi (cfr. esempio n. 72 e 73).

Esempio n. 72 (Prog2_Testo7 – A/Add)

TO: [...] (z. B. bei Verkehrsschildern) oder die Ersetzung eines Gedichtes durch Musik (z. B. Mendelssohns rein instrumentale Konzertouvertüre Meeresstille und glückliche Fahrt, op. 27, über die beiden gleichnamigen Gedichte von Goethe) [...].

TA: [...] (ad esempio i segnali stradali) o la sostituzione di una poesia con la musica (ad esempio l'ouverture da concerto puramente strumentale di Mendelssohn Meeresstille). (ad esempio i segnali stradali) o la sostituzione di una poesia con la musica (ad esempio l'ouverture da concerto puramente strumentale di Mendelssohn Meeresstille und glückliche Fahrt, op. 27, sulle due omonime poesie di Goethe) [...]

PE: [...] (ad esempio i segnali stradali) o la sostituzione di una poesia con la musica (ad esempio l'ouverture da concerto puramente strumentale di Mendelssohn Meeresstille und glückliche Fahrt, op. 27, sulle due omonime poesie di Goethe) [...]

Esempio n. 73 (ProgFT_Testo2 – A/Add)

TO: [...] kann [...] als [...] Reduktionsvarietät des Deutschen [...] beschrieben werden.

TAFT: [...] può essere descritta come [...] come la varietà ridotta del tedesco [...]

PEFT: [...] può essere descritta come [...] la varietà ridotta del tedesco [...]

I due esempi summenzionati mostrano dunque due comportamenti distinti: DeepL (esempio n. 72) tende a produrre aggiunte più estese, strutturalmente invasive, mentre il modello *fine-tuned* commette errori più circoscritti, limitati a singole parole. In entrambi i casi si tratta di aggiunte superflue di interi segmenti o semplici parole.

Tra le categorie che registrano meno errori nel testo generato dal modello *fine-tuned* si individuano le seguenti: F/Sty/Idiom, F/Coh/CoRef, A/MisTra/Gend, F/Coh/Cit e A/MWE/Komp. La categoria F/Sty/Idiom, per la quale si individuano 3 errori nel testo del modello *fine-tuned* e 2 in quello di DeepL, indica le problematiche relative alla traduzione di espressioni idiomatiche. La categoria F/Coh/CoRef segnala, invece, la mancata corrispondenza tra entità citate in frasi diverse e registra 3 errori nel testo del modello *fine-tuned* e 4 in quello di DeepL. L'esempio seguente, tratto dal testo prodotto dal testo di DeepL, illustra tale fenomeno.

Esempio n. 74 (Prog2_Testo7 – F/Coh/CoRef)

TO: [...] wird [...] wieder an die Begriffe “schriftlich” und “Text” gebunden.

TA: [...] è nuovamente collegato ai concetti “scritto” e “testo”.

PE: [...] è nuovamente collegato ai concetti “scritta” e “testo”.

Nell'esempio n. 74, DeepL traduce “schriftlich” con “scritto”, trascurando che tale lessema era già stato introdotto in precedenza come parte di una citazione che riportava l'aggettivo riferito al

lessema “forma”, dunque come “scritta”. Pertanto, nella traduzione deve essere richiamata la forma “scritta” per mantenere la coerenza con l’uso precedente e preservare la relazione corretta tra i referenti.

La categoria A/MisTra/Gend, invece, riguarda gli errori di mancata applicazione del linguaggio inclusivo. Per tale categoria si registra lo stesso numero di occorrenze (2) nei due testi. Tra le categorie più rilevanti per la traduzione del linguaggio scientifico e del genere testuale dell’articolo scientifico rientra anche la categoria F/Coh/Cit, che indica la strutturazione incoerente dei riferimenti bibliografici. Questa categoria presenta una sola occorrenza nel testo generato dal modello *fine-tuned* e 3 nel testo di DeepL, segnalando una gestione più critica dei riferimenti bibliografici da parte del sistema generico. La categoria A/MWE/Komp, relativa alla resa errata di aggettivi composti, presenta una singola occorrenza nel testo del modello *fine-tuned* e nessuna in quello di DeepL.

Infine, si individua la categoria A/Mech/Capit, che concerne il trasferimento errato della lettera maiuscola e registra una sola occorrenza nel testo tradotto da DeepL e nessuna occorrenza nel testo del sistema adattato.

Per riassumere, le differenze più significative riguardano le categorie di F/Coh/InPu (uso incoerente della punteggiatura), F/Lex/LexCho (scelta lessicale non adeguata all’ambito e al genere testuale), A/MisTra/SemUn (traduzione semanticamente errata) e F/Sty/Disfl (traduzione non scorrevole), per le quali il sistema *fine-tuned* registra numeri considerevolmente più elevati di occorrenze rispetto a DeepL. Come già osservato per il testo precedente, anche in questo caso il modello restituisce risultati migliori in riferimento ad alcuni aspetti tipici del linguaggio scientifico, quali la realizzazione delle nominalizzazioni (F/SLP/Nomin), la strutturazione coerente dei riferimenti bibliografici (F/Coh/Cit) e l’uso delle forme impersonali (F/SLP/ForImp). Anche la gestione della terminologia appare in diversi casi più accurata, soprattutto per quanto riguarda la resa dei termini specialistici (A/Termi/SemUn). Permangono, tuttavia, categorie particolarmente problematiche per il modello adattato. Per alcuni fenomeni il numero di errori è analogo nei due sistemi (ad esempio G/Orth/Punct, ossia errori di punteggiatura, F/Sty/Idiom, traduzione errata di espressioni idiomatiche e A/MisTra/Gend per le problematiche nell’uso del linguaggio inclusivo), mentre per altri il modello *fine-tuned* mostra risultati sensibilmente peggiori (ad esempio G/Orth/Spel, cioè errori di ortografia, A/Omis, errori di omissione e A/Mech/CharFor, il trasferimento errato della formattazione). Nonostante criticità individuate, l’insieme dei risultati ottenuti risulta positivo e mostra che il modello adattato costituisce una base solida per l’implementazione di successive ottimizzazioni volte a ottenere una qualità migliore della traduzione del linguaggio scientifico dal tedesco all’italiano in ambito linguistico.

8.2.2 Analisi dell'*effort* temporale e tecnico

Il *post-editing* del testo prodotto dal modello *fine-tuned* è stato svolto in sette sessioni, con durate di 00:39:37, 02:21:54, 01:33:38, 02:31:20, 02:33:27, 03:05:16, 03:11:24, per un totale complessivo di 15:56:36. Per il testo tradotto da DeepL invece il totale delle sessioni ammonta a 10:57:55.

Dall'analisi congiunta delle durate e dell'estensione del testo preso in esame emerge chiaramente che la sola lunghezza del testo (8.484 *tokens*) non costituisce un indicatore affidabile dell'*effort* richiesto per la correzione. Pur trattandosi dello stesso testo di partenza e quindi dello stesso numero di *tokens*, i tempi richiesti per la correzione differiscono in misura sostanziale tra le due traduzioni, suggerendo che l'*effort* non sia determinato tanto dall'estensione del testo quanto piuttosto dalla qualità dell'output generato.

Questa osservazione viene confermata dal confronto tra il numero di errori annotati e la durata del *post-editing*: il testo del modello *fine-tuned* presenta 1.141 errori, mentre il testo di DeepL 794. La maggiore densità di interventi necessaria nel primo caso spiega il tempo di correzione più consistente. Tuttavia, come già sottolineato, non è soltanto la quantità di errori a influire sull'*effort*: la natura degli errori, il loro peso sul piano semantico e stilistico e la loro distribuzione interna rappresentano fattori altrettanto rilevanti.

Prendendo in considerazione il rapporto tra numero di *tokens* e tempo dedicato al *post-editing*, la produttività stimata del *post-editor* risulta di circa 530 *tokens*/ora per il modello *fine-tuned* e di 775 *tokens*/ora per DeepL. Ciò evidenzia come l'output di DeepL risulti più scorrevole nella fase di *post-editing*, permettendo di completare la correzione degli errori in tempi più ridotti rispetto al modello adattato. L'output del modello *fine-tuned*, invece, si caratterizza presumibilmente per una maggiore variabilità qualitativa, che determina tempi di intervento più estesi e una revisione più impegnativa.

Nel complesso, i risultati ribadiscono come la qualità dell'output della TA rappresenti il principale fattore determinante del *post-editing effort*. La mera estensione del testo o il conteggio degli errori non bastano a misurare l'*effort* richiesto, poiché risultano decisivi soprattutto la tipologia delle criticità riscontrate, la loro rilevanza sull'adeguatezza della traduzione e la qualità linguistica complessiva dell'output originario, che incidono in misura sostanziale sulla durata e sulla complessità del processo di correzione.

Dal report di Post-Edit Compare risulta che il testo generato dal sistema *fine-tuned* è composto da 276 segmenti per un totale di 6.728 parole. La quasi totalità, pari a 247 segmenti, proviene dal sistema di TA, mentre 29 segmenti sono stati tradotti manualmente in assenza di una proposta automatica. La traduzione ottenuta tramite DeepL presenta invece 358 segmenti, pur

componendosi dello stesso totale di parole (6.728). La differenza nel numero di segmenti è probabilmente riconducibile a una diversa segmentazione operata dai due sistemi di TA. Nel caso di DeepL, 357 segmenti provengono dalla TA e soltanto uno è stato inserito manualmente. Anche per questo testo non si riscontra alcun contributo della memoria di traduzione. Come già spiegato, ciò dipende dal funzionamento di Post-Edit Compare: lo strumento opera a livello di progetto di traduzione e richiede una versione di partenza unica per tutti i testi, salvata prima dell'avvio del *post-editing*, quando la memoria era ancora vuota. Di conseguenza, analizza esclusivamente tale versione e non ricalcola eventuali *matches* aggiunti durante le successive fasi di correzione. Pertanto, tutto il materiale considerato coincide integralmente con l'output della TA.

La Tabella 16 mostra la distribuzione dei segmenti e delle parole nelle differenti fasce di PEMp. Per ciascuna fascia, le indicazioni di segmenti, parole e percentuali sono organizzate in due sezioni parallele, dedicate rispettivamente al modello *fine-tuned* e a DeepL, consentendo così un confronto diretto e immediato della qualità traduttiva dei due sistemi.

Fasce di PEMp	Segmenti		Parole		Percentuali	
Sistema di TA	Modello <i>fine-tuned</i>	DeepL	Modello <i>fine-tuned</i>	DeepL	Modello <i>fine-tuned</i>	DeepL
100%	19	121	417	1.831	6,27%	27,21%
95-99%	21	16	490	430	7,37%	6,39%
85-94%	54	32	1417	849	21,32%	12,62%
75-84%	65	20	1933	597	29,08%	8,87%
50-74%	48	25	1173	608	17,65%	9,04%
New (0-49%)	57	144	1217	2.413	18,31%	35,87%

Tabella 14: Distribuzione comparative delle fasce di PEMp per testo del modello *fine-tuned* n. 2

L'osservazione della distribuzione dei segmenti nelle differenti fasce di PEMp per il modello *fine-tuned* e per DeepL rivela comportamenti qualitativi sensibilmente diversi. I due sistemi mostrano entrambi una certa irregolarità nei risultati, ma le modalità con cui tale variabilità si manifesta divergono in modo evidente.

Nella fascia 100%, che indica assenza totale di interventi, il divario tra i due sistemi è particolarmente marcato: il modello *fine-tuned* raggiunge soltanto il 6,27% delle parole (19 segmenti), mentre DeepL presenta il 27,21% (121 segmenti). Questo dato suggerisce che DeepL è in grado di generare un numero significativamente più elevato di segmenti immediatamente pubblicabili, senza necessità di apportare modifiche.

Nella fascia 95-99%, relativa agli interventi minimi, i rapporti si riequilibrano: il modello *fine-tuned* totalizza il 7,37% delle parole (21 segmenti), mentre DeepL registra il 6,39% (16

segmenti). Sebbene i valori differiscano di poco, si osserva una lieve inversione rispetto alla fascia precedente, con il sistema adattato che in questa fascia produce una quantità leggermente superiore di segmenti richiedenti modifiche minime.

La fascia 85-94% risulta rilevante per entrambi i sistemi, ma con incidenze diverse: il modello *fine-tuned* concentra qui il 21,32% delle parole (54 segmenti), a fronte del 12,62% (32 segmenti) di DeepL. Questo indica che la testo generato dal sistema *fine-tuned* richiede per oltre un quinto degli interventi di lieve entità. DeepL, invece, presenta una minore incidenza in questa categoria.

La fascia 75-84% evidenzia ulteriormente questa divergenza. Il modello adattato registra il valore più elevato dell'intera distribuzione, con il 29,08% delle parole (65 segmenti), mentre DeepL si limita all'8,87% (20 segmenti). Ciò suggerisce che il modello *fine-tuned* produce numerosi segmenti che necessitano interventi di media entità, evidenziando una qualità più discontinua. La tendenza si conferma anche nella fascia 50-74%, in cui il sistema adattato raggiunge il 17,65% (48 segmenti), contro il 9,04% (25 segmenti) di DeepL.

Infine, la fascia *New* (0-49%) presenta una dinamica inversa: DeepL presenta il valore più elevato dell'intera distribuzione, con il 35,87% delle parole (144 segmenti), mentre il modello *fine-tuned* si attesta al 18,31% (57 segmenti). Questo dato indica che, nei casi in cui DeepL non riesce a generare una proposta di traduzione adeguata, l'errore tende a essere più consistente, richiedendo una rielaborazione integrale o quasi completa del segmento.

Complessivamente, questi risultati delineano due profili qualitativi distinti già rilevati in precedenza. DeepL si distingue nettamente nelle fasce ad alta qualità, offrendo un numero maggiore di segmenti già utilizzabili o che richiedono interventi estremamente contenuti. Tuttavia, quando tale sistema produce errori, questi risultano più gravi, come dimostra la percentuale registrata per la fascia *New*. Il modello *fine-tuned*, invece, mostra una qualità più distribuita nelle fasce intermedie: genera errori meno drastici, collocando la maggior parte del proprio output in fasce che richiedono un intervento umano di entità media ma non totale (fasce 85-94% e 75-84%). Pertanto, DeepL tende a polarizzare l'output tra segmenti molto validi e segmenti molto problematici, mentre il modello *fine-tuned* presenta una qualità più costante, seppur mediamente inferiore.

8.3 Testo per il progetto n.3 (6.120 tokens)

Corrispondente al testo n. 8 del progetto Interculture journal

Stang, Alexandra. 2022. «Theoretische Reflexionen und Konturen einer relational orientierten Wissenskommunikation. Überlegungen für eine künftige Forschungsagenda». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 21 (35): 47–64. <https://doi.org/10.24403/jp.1295924>.

8.3.1 Analisi degli errori

Categoria di errore	Numero di annotazioni nel modello fine-tuned	Numero di annotazioni in DeepL
F/Coh/InPu	151	118
A/MisTra/SemUn	97	57
F/Sty/Disfl	76	78
F/Lex/LexCho	74	71
A/Mech/CharFor	41	2
F/Coh/Cit	33	33
A/MisTra/WSD	22	9
A/Omis	20	22
A/MisTra/Part	16	4
A/MisOrd/WO	12	9
F/Sty/Reg	11	7
A/MisTra/POS	9	6
A/Mech/Punct	9	13
F/Coh/CoRef	9	4
A/Termi/SemUn	9	9
A/SouEr	9	7
A/Add	8	2
A/ForLa	8	10
G/Orth/Punct	7	2
A/MisTra/WroPr	7	2
G/WF/Agree	7	4
A/DNT	6	9
G/Orth/Spel	5	3
G/Orth/WhiSpa	5	4
F/Lex/Fue	4	4
A/Untra	3	25
A/MWE/Ell	3	2
A/MisOrd/MW	3	9
A/MWE/Komp	3	2
F/SLP/Nomin	3	4
G/WF/VF	3	-
F/De/CharFor	2	2
G/Orth/Capit	2	7
A/Oth	2	1
A/MisExpla	2	2
F/De/Layo	1	2
F/Sty/Idiom	1	2
A/Mech/Capit	1	-
F/SLP/ForImp	1	-
A/Termi/DNT	-	3

Tabella 15: Analisi quantitativa comparativa degli errori per il testo del modello *fine-tuned* n. 3

Nel corso dell'analisi del testo tradotto con il sistema *fine-tuned* si riscontrano 366 errori di *Fluency*, 290 errori di *Accuracy* e 29 errori di *Grammar*. Il totale degli errori annotati ammonta a 685. La maggior parte delle annotazioni riguarda la macrocategoria di *Fluency*, che da sola supera la metà del totale. Tale prevalenza è attribuibile in larga misura alla categoria F/Coh/InPu, che conta 151 occorrenze e concerne l'uso incoerente di segni ortografici, quali l'apostrofo e le virgolette (cfr. esempio n. 1 in 8.1.1 e n. 37 in 8.2.1). Nel testo tradotto con il sistema di TA generico DeepL si riscontrano 325 errori di *Fluency*, 205 di *Accuracy* e 20 di *Grammar*, per un totale di 550 annotazioni. Anche in questo caso la maggior parte degli errori concerne la *Fluency* e supera la metà degli errori complessivamente rilevati, con la sottocategoria F/Coh/InPu che registra 118 occorrenze. Nel confronto tra i due sistemi, il modello *fine-tuned* presenta dunque un numero più elevato di errori in tutte le macrocategorie considerate, un dato che si riflette anche nel totale complessivo delle annotazioni.

In ordine decrescente per numero di occorrenze, nel testo del modello *fine-tuned* seguono le seguenti categorie di errori: A/MisTra/SemUn con 97 errori, F/Sty/Disfl con 76 e F/Lex/LexCho con 74. Anche per il testo tradotto da DeepL queste categorie risultano tra le più rappresentate, seppur con un ordine diverso: F/Sty/Disfl con 78 occorrenze, F/Lex/LexCho con 71 occorrenze e A/MisTra/SemUn con 57 occorrenze. Nel complesso, tali categorie mostrano valori più elevati nel testo prodotto dal modello adattato, fatta eccezione per la categoria F/Sty/Disfl, che presenta due occorrenze in più nel testo di DeepL. Per quanto riguarda gli errori di A/MisTra/SemUn, questi segnalano la mancata corrispondenza semantica tra testo di partenza e testo di arrivo. Di seguito viene riportato un esempio tratto dal testo prodotto dal modello *fine-tuned*.

Esempio n. 75 (ProgFT_Testo3 – A/MisTra/SemUn)

TO: [...], und es haben sich dabei unterschiedliche definitorische Annäherungen entwickelt.

TAFT: [...], sviluppando approcci definitivi diversi.

PEFT: [...], sviluppando approcci di definizione diversi.

Nell'esempio n. 75, l'errore riguarda la resa del lessema "definitorische". Il sistema di TA lo interpreta erroneamente come "definitivi", sostituendo il significato originario con uno non pertinente. L'intervento di *post-editing* ripristina la resa corretta, rendendo il lessema con "di definizione", soluzione coerente con il contenuto e il significato del testo di partenza.

La categoria F/Sty/Disfl concerne invece le problematiche di scorrevolezza e la resa dello stile della lingua di arrivo. L'esempio seguente riguarda un errore annotato nel testo prodotto dal sistema *fine-tuned*, che ha richiesto un intervento esteso.

Esempio n. 76 (ProgFT_Testo3 – F/Sty/Disfl)

TO: [...] nicht zuletzt, weil sich die Aussagen auf kein klar umrissenes Netzwerkareal beziehen [...]

TATF: [...] non da ultimo perché le dichiarazioni si riferiscono a nessun'area di rete chiaramente definita [...]

PEFT: [...] non da ultimo perché le dichiarazioni non si riferiscono a un'area di rete chiaramente definita [...]

Nell'esempio n. 76, l'output del sistema *fine-tuned* presenta una resa poco naturale della negazione, influenzata dalla struttura del testo di partenza, con l'impiego dell'indefinito negativo “nessun” in assenza di negazione verbale (“si riferiscono a nessun'area”). Sebbene la costruzione non sia grammaticalmente scorretta, essa risulta poco naturale nella lingua di arrivo. L'intervento di *post-editing* ristabilisce una formulazione più scorrevole (“non si riferiscono a un'area”), migliorando l'autenticità e l'adeguatezza stilistica del segmento.

La categoria F/Lex/LexCho riguarda l'impiego di espressioni non appropriate all'ambito specialistico e al genere testuale specifico. L'esempio n. 77, tratto dal testo del modello *fine-tuned*, mostra un caso in cui il sistema di TA ha impiegato un lessema non conforme alle convenzioni lessicali e stilistiche del linguaggio scientifico.

Esempio n. 77 (ProgFT_Testo3 – F/Lex/LexCho)

TO: [...] die einerseits in der Lage ist, strukturelle Löcher zu überbrücken [...]

TAFT: [...] capace, da un lato, di superare i buchi strutturali [...]

PEFT: [...] capace, da un lato, di superare le lacune strutturali [...]

In riferimento ai fenomeni linguistici del linguaggio scientifico, nel testo prodotto dal modello *fine-tuned* si registrano le seguenti occorrenze: 3 errori di F/SLP/Nomin, riguardanti il mancato impiego di nominalizzazioni e un errore di F/SLP/ForImp, connesso al mancato uso delle forme impersonali in sostituzione di quelle personali. Nel testo prodotto da DeepL si riscontrano invece 4 errori di F/SLP/Nomin e nessun errore per la categoria F/SLP/ForImp. In riferimento agli errori di F/SLP/Nomin, i valori si differenziano per un solo errore, illustrato nei seguenti esempi.

Esempio n. 78 (ProgFT_Testo3 – F/SLP/Nomin)

TO: in Hinblick der reinen Nützlichkeit zum Erreichen bestimmter Ziele

TAFT: in termini di mera utilità per il raggiungimento di determinati obiettivi

Esempio n. 79 (Prog3_Testo8 – F/SLP/Nomin)

TO: in Hinblick der reinen Nützlichkeit zum Erreichen bestimmter Ziele

TA: in termini di pura utilità per raggiungere determinati obiettivi

PE: in termini di pura utilità per il raggiungimento di determinati obiettivi

L'esempio n. 78 presenta il segmento di partenza con la rispettiva traduzione elaborata dal sistema *fine-tuned*, in cui è stata correttamente realizzata una nominalizzazione. Tale fenomeno non si verifica nell'esempio n. 79, relativo al segmento tradotto da DeepL e successivamente post-editato. Nonostante il modello *fine-tuned* presenti un errore in meno, si può affermare che generalmente entrambe le macchine riscontrano difficoltà nel cogliere la tendenza del linguaggio scientifico all'impiego di sintagmi nominali e, in generale, di uno stile nominale, che conferiscono al testo maggiore oggettività e compattezza stilistica, come mostrano gli esempi n. 80 e n. 81.

Esempio n. 80 (ProgFT_Testo3 – F/SLP/Nomin)

TO: [...] kann eine [...] Kommunikation auch zur Überwindung von Unbestimmtheit und Unsicherheit im Umgang mit Anderen beitragen.

TAFT: [...] una comunicazione [...] può contribuire anche a superare l'indeterminatezza e l'insicurezza nei rapporti con gli altri.

PEFT: [...] una comunicazione [...] può contribuire anche al superamento dell'indeterminatezza e dell'insicurezza nei rapporti con gli altri.

Esempio n. 81 (Prog3_Testo8 – F/SLP/Nomin)

TO: kann [...] zur Überwindung von Unbestimmtheit und Unsicherheit im Umgang mit Anderen beitragen.

TA: [...] può contribuire a superare l'incertezza e l'insicurezza nei rapporti con gli altri

PE: [...] può contribuire al superamento dell'incertezza e dell'insicurezza nei rapporti con gli altri.

Gli esempi n. 80 e n. 81 mostrano casi in cui entrambi i sistemi di TA non hanno realizzato la nominalizzazione, impiegando invece una costruzione verbale esplicita, tipica del linguaggio comune. In fase di *post-editing* si è intervenuto in entrambi i casi reintroducendo il sintagma nominale. Negli esempi si possono notare anche degli errori di F/Coh/InPu relativi all'uso incoerente dell'apostrofo. Di seguito viene riportato un esempio per la categoria F/SLP/ForImp (forma impersonale) tratto dal testo generato dal modello *fine-tuned*.

Esempio n. 82 (ProgFT_Testo3 – F/SLP/ForImp)

TO: Aus einer Organisationsperspektive bietet es sich an, zu überlegen, [...]

TAFT: Da una prospettiva organizzativa, ci si propone di riflettere [...]

PEFT: Da una prospettiva organizzativa, si propone una riflessione [...]

Nell'esempio n. 82 il sistema non riconosce la convenzione stilistica del linguaggio scientifico, che privilegia forme impersonali per mantenere un tono neutro e distaccato. Pertanto, nell'intervento di *post-editing* viene impiegata la costruzione “si propone una riflessione”, che integra una nominalizzazione e rispetta le norme stilistiche del linguaggio scientifico.

Per quanto riguarda il lessico del linguaggio scientifico, sono stati riscontrati 4 errori di F/Lex/Fue, riferibili alla traduzione errata delle *Fügungen* sia nel testo prodotto dal sistema di TA *fine-tuned* che in quello generato da DeepL. È interessante notare che, pur registrando lo stesso numero di occorrenze, ciascun sistema di TA mostra punti di forza e debolezza distinti: in alcuni casi la resa fornita dal sistema *fine-tuned* appare più appropriata (cfr. esempio n. 83), mentre in altri si dimostra più adeguata la traduzione proposta da DeepL (cfr. esempio n. 86). Di seguito (esempi 83-86) si citano i casi confrontando le rese dei due sistemi:

Esempio n. 83 (ProgFT_Testo3 – F/Lex/Fue)

TO: Reinmann-Rothermeier (2001:24) versteht darunter [...]

TAFT: Reinmann-Rothermeier (2001:24) intende [...]

Esempio n. 84 (Prog3_Testo8 – F/Lex/Fue)

TO: Reinmann-Rothermeier (2001:24) versteht darunter [...]

TA: Per Reinmann-Rothermeier (2001:24) questo significa [...]

PE: Reinmann-Rothermeier (2001:24) intende [...]

Negli esempi n. 83 e n. 84, relativi alla *Fügung* “versteht darunter”, il sistema *fine-tuned* propone una resa corretta, che preserva la funzione del segmento di partenza e costituisce una costruzione tipica del linguaggio scientifico, utilizzata per chiarire o circoscrivere concetti specialistici. DeepL, invece, riformula il contenuto, risultando meno aderente alla funzione del segmento di partenza. In questo caso, quindi, la versione *fine-tuned* appare più adeguata al corretto uso del linguaggio scientifico quotidiano.

Esempio n. 85 (ProgFT_Testo3 – F/Lex/Fue)

TO: Eine Schlüsselrolle spielen in diesem Zusammenhang Promotoren [...]

TAFT: Un ruolo chiave gioca in questo contesto i promotori [...]

PEFT: I promotori assumono un ruolo fondamentale in questo contesto [...]

Esempio n. 86 (Prog3_Testo8 – F/Lex/Fue)

TO: Eine Schlüsselrolle spielen in diesem Zusammenhang Promotoren [...]

TA: I promotori svolgono un ruolo fondamentale in questo contesto [...]

L'esempio n. 85 illustra un errore della categoria F/Lex/Fue, in cui il sistema *fine-tuned* traduce letteralmente la *Fügung* tedesca senza adattarla al linguaggio scientifico italiano. Oltre alla resa innaturale, si osserva anche un errore di G/WF/VF: il verbo è coniugato alla terza persona singolare ("gioca"), perché il sistema interpreta erroneamente il sintagma "un ruolo chiave" come soggetto della frase, mentre in italiano, tuttavia, la struttura risulta grammaticalmente scorretta, in quanto il soggetto è "i promotori", che richiederebbe una forma plurale. Nell'esempio n. 86, DeepL propone invece una riformulazione idiomatica e sintatticamente corretta. In particolare, evita la topicalizzazione del complemento oggetto, riportando il soggetto nella posizione canonica all'inizio della frase e spostando l'oggetto in posizione postverbale. Questo comporta una modifica dell'ordine dei componenti della frase rispetto alla struttura tedesca originale, ma rende la frase naturale e conforme allo stile del linguaggio scientifico italiano. Una struttura più letterale, come "Un ruolo chiave viene giocato" o "Un ruolo chiave lo giocano", risulterebbe infatti poco naturale e non conforme allo stile del linguaggio scientifico italiano.

Infine, per quanto attiene lo stile del linguaggio scientifico, si segnalano 11 errori di registro inadeguato (F/Sty/Reg) per il testo generato dal sistema *fine-tuned* e 7 errori nel testo prodotto da DeepL, risultati che segnalano una resa migliore di quest'ultimo (cfr. esempio n. 87).

Esempio n. 87 (ProgFT_Testo3 – F/Sty/Reg)

TO: Wird dies zu wenig berücksichtigt [...]

TAFT: Se si tiene troppo poco conto di questo fatto [...]

PEFT: Se non si considera adeguatamente questo aspetto [...]

L'uso corretto della terminologia costituisce un aspetto fondamentale del linguaggio scientifico: a questo proposito nel testo prodotto dal sistema *fine-tuned* si registrano 9 errori di A/Termi/SemUn, relativi alla traduzione errata di termini, e nessun errore di A/Termi/DNT, categoria impiegata per indicare termini che non necessitano di traduzione. Invece nel testo tradotto da DeepL si verificano 9 errori di A/Termi/SemUn e 3 errori di A/Termi/DNT. Di seguito vengono riportati due esempi per la categoria A/Termi/SemUn e relativi allo stesso segmento: l'esempio n. 88 per il testo del modello adattato e l'esempio n. 89 per il testo di DeepL.

Esempio n. 88 (ProgFT_Testo3 – A/Termi/SemUn)

TO: Aus der Perspektive der Akteur-Netzwerk-Theorie [...]

TAFT: Dal punto di vista dell'attore-rete-teoria [...]

PEFT: Dal punto di vista della teoria dell'attore-rete [...]

Esempio n. 89 (Prog3_Testo8 – A/Termi/SemUn)

TO: Aus der Perspektive der Akteur-Netzwerk-Theorie [...]

TA: Dal punto di vista della teoria degli attori-rete [...]

PE: Dal punto di vista della teoria dell'attore-rete [...]

I sistemi di TA presi in esame commettono lo stesso quantitativo di errori nella traduzione di termini (A/Termi/SemUn), non riuscendo a fornire l'equivalente corretto nella lingua di arrivo. Negli esempi n. 88 e n. 89, entrambi i sistemi traducono in modo impreciso la terminologia "Akteur-Netzwerk-Theorie", richiedendo un intervento di correzione per ripristinare la forma corretta "teoria dell'attore-rete".

Esempio n. 90 (ProgFT_Testo3 – A/Termi/DNT)

TO: [...] hat Manuel Castells (1996) [...] den Begriff der Network Society geprägt.

TAFT: Manuel Castells (1996) ha coniato il concetto di Network Society [...]

Esempio n. 91 (Prog3_Testo8 – A/Termi/DNT)

TO: hat Manuel Castells (1996) [...] den Begriff der Network Society geprägt.

TA: Manuel Castells (1996) ha coniato il termine rete società [...]

PE: Manuel Castells (1996) ha coniato il concetto di Network Society [...]

L'esempio n. 90 mette in evidenza la capacità del sistema *fine-tuned* di riconoscere l'uso consolidato di termini in inglese nella lingua di arrivo. Questo non si verifica nell'esempio n. 91, tratto da DeepL. Dal confronto dei dati per queste due categorie, emerge una gestione migliore della terminologia da parte del sistema *fine-tuned*, in particolare in riferimento ai termini che non necessitano una traduzione.

Un'ulteriore categoria con una frequenza significativa è A/Mech/CharFor (rispetto della formattazione del testo di partenza), che presenta 41 occorrenze nel testo prodotto dal sistema adattato e riguarda gli errori nel trasferimento della formattazione. Nel testo generato da DeepL invece si verificano solo 2 casi. Di seguito viene riportato un esempio tratto dal testo del modello *fine-tuned*.

Esempio n. 92 (ProgFT_Testo3 – A/Mech/CharFor)

TO: Ein zentrales Ziel von Wissenskommunikation ist es [...]

TAFT: Un obiettivo centrale della comunicazione delle conoscenze è [...]

PEFT: Un obiettivo centrale della comunicazione delle conoscenze è [...]

Nel caso dell'esempio n. 92, il segmento originale è interamente in corsivo, mentre nella proposta di traduzione del sistema *fine-tuned* tale formattazione viene applicata solo su alcune parole. Si tratta di un errore di natura tecnica, riconducibile verosimilmente al processo di esportazione o di codifica del testo durante la traduzione automatica, e non a una scelta linguistica del sistema. Di conseguenza, questi casi non influenzano la qualità traduttiva sul piano contenutistico, ma incidono esclusivamente sull'aspetto formale del testo.

Tra le categorie di errori più rilevanti per la traduzione del linguaggio scientifico e del genere testuale dell'articolo scientifico, si individua la categoria F/Coh/Cit, che riguarda la strutturazione incoerente dei riferimenti bibliografici. In entrambi i testi si registrano 33 occorrenze per questa categoria (cfr. esempi 93 e 94).

Esempio n. 93 (ProgFT_Testo3 – F/Coh/Cit)

TO: (Stegbauer & Haussling 2010)

TAFT: (Stegbauer & Haussling 2010)

PEFT: (Stegbauer e Haussling 2010)

Esempio n. 94 (Prog3_Testo8 – F/Coh/Cit)

TO: (Stegbauer & Haussling 2010)

TA: (Stegbauer & Haussling 2010)

PE: (Stegbauer e Haussling 2010)

Nelle proposte di traduzione dei due sistemi viene mantenuto il simbolo “&”, tipico delle convenzioni anglosassoni. L'intervento di *post-editing* sostituisce tale simbolo con la congiunzione “e”, adeguando il riferimento bibliografico alle convenzioni citazionali impiegate nei testi di partenza, ossia *Chicago Manual of Style* (17^a ed.), e, al tempo stesso, alla lingua di arrivo. I risultati indicano dunque che i sistemi di TA incontrano difficoltà analoghe nella gestione dei riferimenti bibliografici.

Proseguendo in ordine decrescente di occorrenze, il testo tradotto mediante il sistema *fine-tuned* presenta le seguenti categorie di errori: A/MisTra/WSD per gli errori di scelta semantica in riferimento a parole polisemiche (22 occorrenze); A/Omis per le omissioni di segmenti (20 occorrenze); e A/MisTra/Part per gli errori riguardanti il trasferimento parziale o impreciso del significato (16 occorrenze). Nel testo tradotto da DeepL si individuano le seguenti distribuzioni: A/MisTra/WSD con 9 errori, A/Omis con 22 e A/MisTra/Part con 4. Emergono, pertanto, differenze consistenti nei valori per le categorie A/MisTra/WSD e A/MisTra/Part, per le quali il testo tradotto dal modello *fine-tuned* registra un'incidenza nettamente superiore. Il testo tradotto da

DeepL, invece, presenta un valore leggermente più elevato per la categoria A/Omis. Si tratta di tipologie di errore di *Accuracy* che possono incidere in maniera significativa sulla trasmissione del contenuto e, di conseguenza, sulla qualità complessiva del testo scientifico prodotto.

Esempio n. 95 (ProgFT_Testo3 – A/MisTra/Part)

TO: Neuere Arbeiten greifen auch die aktuellen Medien- und Internetentwicklungen [...] auf, [...]

TAFT: Lo studio più recente riprende anche gli attuali sviluppi dei media e di internet [...]

PEFT: Gli studi più recenti riprendono anche gli attuali sviluppi dei media e di internet [...]

L'esempio n. 95 mostra un caso appartenente alla categoria A/MisTra/Part, in cui il sistema *fine-tuned* non trasferisce correttamente il numero del sostantivo e dell'aggettivo, né del verbo a questi collegato. Errori di questo tipo possono manifestarsi anche come un trasferimento inaccurato del genere o del tempo verbale. Tali problematiche ricorrono sia nel testo prodotto dal modello *fine-tuned* sia in quello generato da DeepL e impattano la fedeltà semantica al testo di partenza e non la correttezza grammaticale della lingua di arrivo.

Per quanto riguarda gli errori di A/Omis, si osserva che nel testo tradotto da DeepL questi riguardano nella maggior parte dei casi più parole o anche segmenti interi (cfr. esempio n. 96). Invece, nel caso del testo tradotto dal sistema *fine-tuned*, gli errori si manifestano in riferimento a singole parole o parti di parole (cfr. esempio n. 97)

Esempio n. 96 (Prog3_Testo8 – A/Omis)

TO: Heterogenität und Hybridität in Netzwerken aus seiner Sicht am besten gerecht werden kann.

TA: eterogeneità e ibridazione nelle reti.

PE: In quest'ottica, può soddisfare al meglio l'eterogeneità e l'ibridismo nelle reti.

Nell'esempio n. 96, l'omissione dipende da una difficoltà di analisi sintattica delle strutture della lingua di arrivo da parte di DeepL. La frase di partenza è priva di soggetto esplicito, che è sottinteso nella frase precedente. Questa mancanza rende complessa la ricostruzione delle relazioni tra gli elementi da parte del sistema, che tende così a interrompere la traduzione, producendo un segmento incompleto e determinando una perdita di contenuto semantico. Oltre all'omissione, si osserva un errore di A/MisTra/SemUn: la traduzione automatica interpreta il lessema “Hybridität” come un processo (“ibridazione”) piuttosto che come condizione o stato (“ibridismo”), generando una mancata corrispondenza semantica rispetto al testo originale. La proposta di *post-editing* corregge questo trasferimento errato, ripristinando il significato originario.

Esempio n. 97 (ProgFT_Testo3 – A/Omis)

TO: [...] im Kontext gesellschaftlicher Paradigmenverschiebungen [...]

TAFT: [...] nel contesto dei paradigmi [...]

PEFT: [...] nel contesto di cambiamenti di paradigma

Nell'esempio n. 97 del testo del modello *fine-tuned*, l'errore nasce da un'analisi morfologica incompleta: il sistema interpreta solo una parte del composto, omettendo l'elemento rimanente e causando la perdita di una componente informativa rilevante.

Altre categorie di *Accuracy* riscontrate sono A/MisOrd/WO e A/MisTra/POS. Si tratta di categorie che influenzano la comprensibilità generale e la fedeltà al testo di partenza, senza impattare la correttezza grammaticale e formale della lingua di arrivo. Più nello specifico, la categoria A/MisOrd/WO si riferisce alle variazioni del significato determinate dall'ordine delle parole, che possono compromettere la corretta trasmissione del messaggio di partenza nel testo di arrivo. Questa registra 12 occorrenze nel testo generato dal modello *fine-tuned* e 9 in quello generato da DeepL. Di seguito viene presentato un esempio tradotto dal testo del sistema adattato.

Esempio n. 98 (ProgFT_Testo3 – A/MisOrd/WO)

TO: Diese Heterogenität kann einerseits [...] Potenzial bieten, innovative Ideen gemeinsam zu entwickeln, sie kann andererseits auch als Herausforderung betrachtet werden, wenn kein gemeinsames mentales Modell [...] vorhanden ist.

TAFT: Questa eterogeneità può, da un lato, offrire delle potenzialità per [...] sviluppare insieme idee innovative, dall'altro può essere vista come una sfida anche se manca un modello mentale comune [...]

PEFT: Questa eterogeneità può, da un lato, [...] offrire delle potenzialità per sviluppare insieme idee innovative, dall'altro può essere vista anche come una sfida se manca un modello mentale comune [...]

Nell'esempio n. 98, l'avverbio “anche” è posizionato nella TAFT in modo tale da modificare il senso della frase, trasformando la funzione del lessema da particella aggiuntiva a congiunzione concessiva tramite la combinazione con “se”. La frase risulta quindi grammaticalmente corretta, ma il significato e la relazione logica e semantica tra i lessemi viene alterata, compromettendo la fedeltà al testo di partenza. L'intervento di *post-editing* ripristina la posizione originale di “anche”, garantendo una resa adeguata del contenuto e delle relazioni tra i lessemi.

La categoria A/MisTra/POS, invece, segnala l'impiego di una categoria lessicale (*Wortart*) non corrispondente a quella del testo di partenza. Per questa categoria sono stati individuati 9 errori

nel testo prodotto dal sistema adattato e 6 nel testo di DeepL. L'esempio n. 99 riporta un caso per questa categoria individuato nel testo generato da DeepL.

Esempio n. 99 (Prog3_Testo8 – A/MisTra/POS)

TO: [...] „nachhaltig konstruktive und kooperative interkulturelle Zusammenarbeit“ (Bolten 2018:10) [...]

TA: [...] "collaborazione interculturale costruttiva e cooperativa sostenibile" (Bolten 2018:10) [...]

PE: [...] “collaborazione interculturale costruttiva e cooperativa a lungo” (Bolten 2018:10) [...]

Nel segmento di partenza dell'esempio n. 99, il lessema “nachhaltig” ha funzione avverbiale, mentre nella traduzione automatica proposta da DeepL è reso come un aggettivo (“sostenibile”) la cui scelta causa anche una modifica del significato. La proposta di traduzione è quindi grammaticalmente corretta, ma l'uso improprio della categoria lessicale altera la funzione sintattica e semantica del lessema originale, compromettendo la fedeltà al testo di partenza e causando inoltre uno slittamento semantico. La proposta di *post-editing* corregge la resa in modo da ripristinare la funzione avverbiale e il significato originale (“a lungo”).

Un'ulteriore categoria osservata è A/Mech/Punct, che indica gli errori relativi al trasferimento della punteggiatura. Nel testo generato dal sistema *fine-tuned* si registrano 9 occorrenze (cfr. esempio n. 100), mentre nel testo tradotto da DeepL si contano 13 casi (cfr. esempio n. 101), evidenziando una maggiore difficoltà di questo sistema nel trasferimento della punteggiatura.

Esempio n. 100 (ProgFT_Testo3 – A/Mech/Punct)

TO: [...] die „interaktive Wissenskonstruktion durch den Austausch von Ergebnissen der Reflexion“ hervorhebt.

TAFT: [...] sottolinea la "costruzione interattiva della conoscenza attraverso la condivisione dei risultati della riflessione"

PEFT: [...] sottolinea la “costruzione interattiva della conoscenza attraverso la condivisione dei risultati della riflessione”

Nell'esempio n. 100, il modello *fine-tuned* sostituisce le virgolette tipografiche tedesche con una combinazione non corretta di virgolette doppie e singole ("..."). L'errore riguarda dunque un problema di adattamento formale piuttosto che di contenuto, che evidenzia una gestione parziale della punteggiatura nel passaggio da un sistema tipografico all'altro. Si riscontra inoltre un errore di F/Coh/InPu legato all'uso incoerente delle virgolette dritte.

Esempio n. 101 (Prog3_Testo8 – A/Mech/Punct)

TO: [...] gegenüber dem jeweiligen Sachverhalt *mehr* oder *minder* angemessen erscheinen können.

TA: [...] possono sembrare *più* o *meno* appropriate alla rispettiva circostanza.

PE: [...] possono sembrare *più* o *meno* appropriate alla rispettiva circostanza.

Nell'esempio n. 101, tratto dal testo di DeepL, si riscontra invece un caso di aggiunta impropria di punteggiatura, con la presenza di due punti fermi consecutivi al termine della frase. Si tratta di un errore di tipo meccanico e si presenta con maggiore frequenza in DeepL rispetto al sistema *fine-tuned*, suggerendo una minor stabilità nel trasferimento formale tra i due sistemi.

Proseguendo nell'analisi si riscontrano le seguenti categorie: F/Coh/CoRef, A/SouEr e A/Add. In queste categorie di errori si riscontra una tendenza molto chiara: il modello *fine-tuned* tende a registrare più errori rispetto a DeepL. La categoria F/Coh/CoRef riguarda la mancata corrispondenza tra referenti citati in frasi diverse e presenta 9 errori nel testo del modello *fine-tuned* e 4 nel testo di DeepL). Di seguito viene riportato un esempio tratto dal testo del sistema adattato.

Esempio n. 102 (ProgFT_Testo3 – A/SouEr)

TO: Dazu zählen die [...] Plattform „Experience Map“ [...]

TAFT: Tra questi si annovera la piattaforma [...] "Experience Map" [...]

PEFT: Tra queste si annovera la piattaforma [...] "Experience Map" [...]

Nell'esempio n. 102 si osserva un errore di coreferenza: il dimostrativo “questi” non stabilisce una relazione corretta con l'elemento menzionato nella frase precedente a cui fa riferimento (“iniziative”). Tale incoerenza non dà luogo a una frase grammaticalmente scorretta, ma danneggia la coesione testuale e la corretta trasmissione del messaggio di partenza. L'intervento di *post-editing* ristabilisce la continuità referenziale sostituendo “questi” con “queste”.

La categoria A/SouEr segnala invece gli errori presenti nel testo di partenza. Per tale categoria si registrano 9 errori nel testo del sistema adattato e 7 nel testo di DeepL. Gli errori nel testo generato dal modello *fine-tuned* derivano in quattro casi su nove dalla collocazione errata del numero della nota a piè di pagina, errore che si verifica nel file traducibile del testo di partenza e nel testo di arrivo creati da Trados Studio, ma non nel file originale del testo di partenza (cfr. esempio n. 36 in 8.1.1). Gli altri casi riguardano parole scritte in modo errato o punteggiatura superflua o mancante. Errori analoghi si riscontrano nel testo tradotto da DeepL (cfr. esempi n. 103 e n. 104).

Esempio n. 103 (ProgFT_Testo3 – A/SouEr)

TO: die Welt ein stückweit zu verändern (Bergmann 2007:12)

TAFT: a cambiare un po' il mondo (Bergmann 2007:12)

PEFT: a cambiare un po' il mondo (Bergmann 2007:12).

Esempio n. 104 (Prog3_Testo8 – A/SouEr)

TO: [...] die Welt ein stückweit zu verändern (Bergmann 2007:12)

TA: [...] a cambiare un po' il mondo (Bergmann 2007:12)

PE: [...] a cambiare un po' il mondo (Bergmann 2007:12).

In entrambi gli esempi summenzionati si constata l'assenza del punto fermo alla fine della frase, un errore presente già nel testo di partenza che viene trasmesso dal sistema di TA nella sua proposta di traduzione.

Infine, la categoria A/Add riguarda i segmenti aggiunti, registrando 8 errori nel testo del sistema adattato e 2 nel testo di DeepL, di seguito vengono mostrati un esempio (105) per il testo del modello *fine-tuned* e uno (106) per il testo tradotto da DeepL.

Esempio n. 105 (ProgFT_Testo3 – A/Add)

TO: [...] auch wenn es andere Perspektiven zu einem bestimmten Sachverhalt gibt.

TAFT: [...] anche se vi sono altre prospettive su un dato dato dato fatto.

PEFT: [...] anche se vi sono altre prospettive su un dato fatto.

Esempio n. 106 (Prog3_Testo8 – A/Add)

TO: [...] (Rosa / Strecker / Kottmann 2018:235).

TA: [...] (Rosa / Strecker / Kottmann 2018:235). (Rosa / Strecker / Kottmann 2018:235)

PE: [...] (Rosa / Strecker / Kottmann 2018:235).

In entrambi i casi si tratta di errori riconducibili a fenomeni di sovragerazione (*overgeneration*), ossia alla produzione da parte del sistema di TA di materiale linguistico superfluo rispetto al testo di partenza, in particolare ripetizioni o aggiunte di parole o segmenti, che alterano la struttura sintattica e non riflettono il significato originale. Tendenzialmente DeepL replica segmenti più lunghi, come illustrato nell'esempio n. 106; il modello *fine-tuned* invece tende a replicare o aggiungere singoli elementi.

La categoria A/ForLa presenta una tendenza in contrasto rispetto alle ultime categorie: si individuano 8 errori nel testo generato dal modello *fine-tuned*, a fronte di 10 nel testo tradotto da DeepL, che quindi registra un valore più elevato. Nel testo del sistema adattato, gli errori di A/ForLa sono legati all'uso improprio della lingua inglese in sostituzione della lingua di arrivo richiesta e, in alcuni casi, anche all'uso di parole inesistenti in italiano (cfr. esempio n. 107). Nel

testo tradotto da DeepL, invece, gli errori di questo tipo riguardano esclusivamente l'uso non richiesto della lingua inglese.

Esempio n. 107 (ProgFT_Testo3 – A/ForLa)

TO: Das Thematisieren einer gemeinsamen Basis kann die Kohäsion untereinander stärken, [...]

TAFT: Temizzare una base condivisa può rafforzare la coesione tra di loro [...]

PEFT: La tematizzazione di una base condivisa può rafforzare la coesione tra di loro [...]

Nell'esempio n. 107 il modello *fine-tuned* produce il verbo “temizzare”, forma inesistente nella lingua italiana e derivata impropriamente dal tedesco “thematisieren”. In fase di *post-editing*, la riformulazione in “La tematizzazione di una base condivisa...” corregge questa deviazione lessicale e morfologica e introduce un sintagma nominale, determinando anche l'annotazione di un errore di F/SLP/Nomin. Tale nominalizzazione consente di restituire una struttura più conforme allo stile del linguaggio scientifico, che predilige costrutti nominali ad alta astrattezza e compattezza.

La categoria A/MisTra/WroPr, riguardante gli errori di preposizioni, sintagmi preposizionali o congiunzioni, registra 7 occorrenze nel testo prodotto dal modello *fine-tuned* e 2 nel testo tradotto da DeepL, evidenziando dunque maggiori criticità nel sistema adattato. Si tratta di errori che non compromettono la correttezza grammaticale della frase in lingua di arrivo, ma alterano le relazioni semantiche tra gli elementi, mettendo a rischio la fedeltà al significato del testo di partenza (cfr. esempio n. 24 in 8.1.1).

Per altre categorie di errori si osserva una tendenza opposta, con valori più elevati nel testo tradotto da DeepL: la categoria A/DNT, relativa ai segmenti la cui traduzione non è necessaria, registra 6 errori nel testo del modello *fine-tuned* e 9 nel testo di DeepL; la categoria A/Untra, che segnala i segmenti non tradotti, registra 3 errori nel testo del sistema adattato e 25 errori nel testo di DeepL; infine, la categoria A/MisOrd/MW, relativa alla collocazione sbagliata di un elemento di un composto, presenta 3 errori nel testo del modello *fine-tuned* e 9 nel testo di DeepL. Tali categorie di *Accuracy* influenzano in modo determinante la trasmissione corretta del contenuto e la qualità complessiva del testo scientifico, senza tuttavia danneggiare la correttezza formale della lingua di arrivo. Di particolare rilievo è la differenza dei valori per la categoria A/Untra: nel testo prodotto dal sistema *fine-tuned*, il fenomeno interessa prevalentemente singoli lessemi (cfr. esempio n. 108), mentre nel testo di DeepL può estendersi fino a coinvolgere intere frasi (cfr. esempio n. 109), evidenziando modalità di errore differenti tra i due sistemi.

Esempio n. 108 (ProgFT_Testo3 – A/Untra)

TO: Sie wirken interdependent [...]

TAFT: Fanno l'interdependent, [...]

PEFT: Operano in modo interdipendente [...]

Esempio n. 109 (Prog3_Testo8 – A/Untra)

TO: Im weitesten Sinne können kommunikative Handlungen einen Beitrag dazu leisten, die Welt ein stückweit zu verändern (Bergmann 2007:12)

TA: Im weitesten Sinne können kommunikative Handlungen einen Beitrag dazu leisten, die Welt ein stückweit zu verändern (Bergmann 2007:12)

PE: In senso lato, le azioni comunicative possono contribuire a cambiare in parte il mondo (Bergmann 2007:12)

Gli esempi n. 110 e n. 111 mostrano come gli errori di A/DNT (traduzione non necessaria) si manifestino in modo differente nei due testi. Nel sistema *fine-tuned*, gli errori di questa categoria rappresentano soprattutto interferenze lessicali parziali: il modello tende a lasciare inalterato il segmento in lingua originale, ma traduce occasionalmente singole parole per adattare al contesto italiano, generando così segmenti ibridi, come mostra l'esempio n. 110 che riporta una citazione in inglese.

Esempio n. 110 (ProgFT_Testo3 – A/DNT)

TO: „to co-develop a research agenda for advancing a relational paradigm within sustainability research, practice, and education based on relational ways of being, knowing, and acting.“

TAFT: “to co-develop a research agenda for advancing a relazionale paradigm within sustainability research, practice, and education based on relational ways of being, knowing, and acting.”

PEFT: “to co-develop a research agenda for advancing a relational paradigm within sustainability research, practice, and education based on relational ways of being, knowing, and acting.”

Nell'esempio n. 111, invece, gli errori di tale categoria tendono a riguardare l'intero segmento destinato a restare in lingua originale. Ciò si verifica perché il sistema DeepL non riconosce le sequenze citazionali in una lingua diversa da quella di partenza e, di conseguenza, considera ogni porzione del testo come traducibile.

Esempio n. 111 (Prog3_Testo8 – A/DNT)

TO: „to co-develop a research agenda for advancing a relational paradigm within sustainability research, practice, and education based on relational ways of being, knowing, and acting.“

TA: "co-sviluppare un'agenda di ricerca per far avanzare un paradigma relazionale all'interno della ricerca, della pratica e dell'educazione alla sostenibilità, basato su modi relazionali di essere, conoscere e agire".

PE: "to co-develop a research agenda for advancing a relational paradigm within sustainability research, practice, and education based on relational ways of being, knowing, and acting."

La macrocategoria *Grammar* comprende errori relativi ad aspetti della grammatica e dell'ortografia, quali l'accordo tra parole (G/WF/Agree), la punteggiatura (G/Orth/Punct), la spaziatura (G/Orth/WhiSpa), l'ortografia (G/Orth/Spel), l'uso delle lettere maiuscole (G/Orth/Capit) e i tempi e le forme verbali (G/WF/VF). Nel testo prodotto dal sistema *fine-tuned*, tali errori risultano più numerosi. Si osservano le seguenti distribuzioni: G/WF/Agree e G/Orth/Punct con 7 errori ciascuna, G/Orth/WhiSpa e G/Orth/Spel con 5, G/WF/VF con 3 e G/Orth/Capit con 2. Nel testo generato da DeepL, invece, la frequenza complessiva degli errori grammaticali è inferiore e la loro distribuzione presenta un profilo diverso: si contano 7 errori per G/Orth/Capit, 4 per G/WF/Agree e G/Orth/WhiSpa, 3 per G/Orth/Spel, 2 per G/Orth/Punct e nessuna occorrenza per G/WF/VF. Nel complesso, dunque, il sistema *fine-tuned* mostra una maggiore incidenza di errori grammaticali e ortografici (cfr. esempio n. 112), mentre DeepL evidenzia un numero complessivamente più contenuto, fatta eccezione per gli errori di G/Orth/Capit, relativi all'uso scorretto delle lettere maiuscole (cfr. esempio n. 113).

Esempio n. 112 (ProgFT_Testo3 – G/WF/VF)

TO: [...] auch eine der jeweiligen Zielgruppen angemessene Präsentation zu wählen (Kade 2007:214), die es im Anschluss ermöglichen [...] in den Austausch und Diskurs über die Inhalte zu treten.

TAFT: [...] la scelta di una presentazione adeguata ai rispettivi destinatari (Kade 2007:214), che permette poi allo scambio e il dibattito sui contenuti [...].

PEFT: [...] anche la scelta di una presentazione adeguata ai rispettivi destinatari (Kade 2007:214), che permetta poi lo scambio e il dibattito sui contenuti [...].

Nell'esempio sopracitato, l'errore riguarda l'uso improprio della forma verbale (G/WF/VF): il sistema *fine-tuned* impiega l'indicativo presente ("che permette") al posto del congiuntivo presente ("che permetta"), più appropriato nel contesto sintattico preso in esame. Tale problematica non concerne il contenuto del segmento e, pertanto, non incide sull'*Accuracy*, ma dà luogo a una frase grammaticalmente scorretta. Questo caso evidenzia una difficoltà del modello adattato nel gestire i vincoli di reggenza verbale all'interno della frase e, di conseguenza, nel selezionare il modo verbale

appropriato. Si segnala inoltre la presenza di errori nel testo di partenza (A/SouEr): nella proposizione relativa introdotta da “die” manca la virgola di separazione prima della proposizione infinitiva e il verbo “ermöglichen” è al plurale, sebbene il pronome relativo “die” si riferisca al sostantivo singolare “Präsentation”. Tale incongruenza grammaticale non viene riprodotta nella traduzione automatica, che utilizza correttamente una forma verbale al singolare, ma può aver influito sulla resa del segmento da parte del sistema.

Esempio n. 113 (Prog3_Testo8 – G/Orth/Capit)

TO: Überlegungen zu Perspektive [...]

TA: Riflessioni sulle Prospettive [...]

PE: Riflessioni sulle prospettive [...]

Nell'esempio n. 113, tratto dal testo prodotto da DeepL, si osserva un esempio per la categoria G/Orth/Capit, consistente nell'uso improprio della lettera maiuscola nella parola “Prospettive”. Si tratta di un errore di natura ortografica e non semantica, privo di effetti sulla comprensibilità del testo, ma indicativo di un'attenzione minore agli aspetti formali della traduzione.

Nel testo generato dal sistema *fine-tuned*, le categorie relative ai composti registrano 3 occorrenze ciascuna per le categorie A/MWE/Ell e A/MWE/Komp; nel testo generato da DeepL 2 occorrenze per ciascuna categoria. La categoria A/MWE/Ell riguarda gli aspetti problematici legati alla traduzione di composti con ellissi di elementi ripetuti (cfr. esempio n. 114). Invece la categoria A/MWE/Komp riguarda gli errori nella traduzione di aggettivi composti (cfr. esempio n. 115). I dati registrati evidenziano quindi una gestione più problematica di questi elementi da parte del modello *fine-tuned*.

Esempio n. 114 (ProgFT_Testo3 – A/MWE/Ell)

TO: [...] medien- und kommunikationstechnologischer Entwicklungen [...]

TAFT: [...] degli sviluppi di media e di comunicazione tecnologica [...]

PEFT: [...] degli sviluppi della tecnologia dei media e della comunicazione [...]

Nell'esempio n. 114, viene illustrato un caso per la categoria A/MWE/Ell, che riguarda la traduzione di composti tedeschi contenenti un'ellissi marcata tramite *Ergänzungsstrich*. In questo caso il sistema di TA non riconosce la struttura ellittica del composto tedesco e interpreta i due elementi come sintagmi autonomi (“media” e “comunicazione tecnologica”), perdendo la relazione morfologica e semantica interna al composto. L'intervento di *post-editing* ricostruisce la struttura complessa, rendendo esplicita l'unità concettuale sottostante e preservando il contenuto semantico che la TA aveva omissso.

Esempio n. 115 (ProgFT_Testo3 – A/MWE/Komp)

TO: [...] die radikal-konstruktivistische Denkweise [...]

TAFT: [...] il pensiero radicale-costruttivista [...]

PEFT: [...] il pensiero costruttivista-radicale [...]

Nell'esempio n. 115 emerge un caso rappresentativo della categoria A/MWE/Komp, che riguarda la resa dei composti aggettivali tedeschi. Nel sintagma “radikal-konstruktivistische”, “konstruktivistisch” costituisce il determinato, mentre “radikal” funge da determinante, specificandone l'orientamento. La traduzione automatica (“pensiero radicale-costruttivista”) non rispetta questa gerarchia, alterando il significato del composto. Pertanto, nella proposta di *post-editing* è stato ristabilito il rapporto corretto tra determinante e determinato (“pensiero costruttivista-radicale”), adeguando la struttura del composto alle convenzioni dell'italiano.

Si rilevano lo stesso numero di occorrenze nei due testi tradotti automaticamente per le categorie F/De/CharFor e A/MisExpl, con 2 occorrenze ciascuna. La categoria F/De/CharFor indica gli errori relativi al mancato rispetto delle convenzioni tipografiche del testo di arrivo, che richiedono una formattazione diversa. Invece, la categoria A/MisExpl segnala i casi di inadeguatezza dei contenuti per il destinatario del testo di arrivo. Gli errori registrati riguardano le stesse problematiche, di cui di seguito viene fornito un esempio.

Esempio n. 116 (ProgFT_Testo3 – A/MisExpl + F/De/CharFor)

TO: [...] kann es hilfreich sein, sich für Umsetzungsüberlegungen an der Sandbergmetapher [...] zu orientieren [...]

TAFT: [...] può essere utile per le riflessioni di implementazione orientarsi alla metafora di Sandberg [...]

PEFT: [...] per le riflessioni di implementazione può essere utile orientarsi alla metafora della montagna di sabbia (Sandbergmetapher) [...]

Esempio n. 117 (Prog3_Testo8 – A/MisExpl + F/De/CharFor)

TO: [...] sich für Umsetzungsüberlegungen an der Sandbergmetapher [...] zu orientieren [...]

TA: [...] può essere utile per le considerazioni sull'implementazione orientarsi alla metafora di Sandberg [...]

PE: [...] per le considerazioni sull'implementazione può essere utile orientarsi alla metafora della montagna di sabbia (Sandbergmetapher) [...]

In entrambi gli esempi si riscontrano simultaneamente un errore di A/MisExpl e uno di F/De/CharFor. L'errore di A/MisExpl consiste nella resa da parte dei sistemi di TA del lessema

“Sandbergmetapher” come “metafora di Sandberg”, una traduzione semanticamente errata che interpreta il contenuto concettuale del termine originale come un nome proprio. L’intervento di *post-editing* ripristina il valore concettuale originario del termine mediante la resa “metafora della montagna di sabbia”, necessaria a rendere comprensibile il riferimento culturale e concettuale nel testo di arrivo. L’errore F/De/CharFor riguarda invece la mancata applicazione delle convenzioni tipografiche italiane: il termine tedesco *Sandbergmetapher*, mantenuto nel testo per ragioni terminologiche, deve essere riportato in corsivo per segnalarne lo status di elemento lessicale straniero. Si osserva inoltre un errore di F/Sty/Disfl riguardante l’espressione “può essere utile”, la cui posizione nella traduzione automatica genera una struttura poco scorrevole.

Tra le categorie che registrano meno errori si individuano F/De/Layo e F/Sty/Idiom. La categoria F/De/Layo comprende gli errori relativi alla strutturazione dei paragrafi e del testo di arrivo e registra un errore nel testo generato dal sistema *fine-tuned* e 2 errori nel testo di DeepL. Gli stessi valori sono stati individuati per la categoria F/Sty/Idiom, riguardante la traduzione errata di espressioni idiomatiche. In entrambi i casi, si tratta di errori che incidono sulla scorrevolezza e sulla strutturazione formale del testo, influenzandone quindi anche la qualità complessiva. I dati indicano che, pur in misura limitata, la gestione di queste problematiche risulta leggermente più efficace nel sistema *fine-tuned*.

Per la categoria A/Oth si riscontrano 2 errori nel testo tradotto dal sistema *fine-tuned*, mentre un solo errore nel testo di DeepL. Tale categoria di errori segnala i casi problematici non riconducibili ad altre categorie di errori. Un errore di A/Oth riguarda in un caso l’uso della parola “tedesco” accanto ad “Abstract”: la traduzione è fedele, ma non adeguata al contesto del testo di arrivo, dove l’abstract è in italiano, come il contributo in traduzione. Questa problematica viene riscontrata anche nel testo tradotto da DeepL, come mostra l’esempio n. 118. L’altro caso è attribuibile al trasferimento errato di un nome proprio (cfr. esempio n. 119).

Esempio n. 118 (Prog3_Testo8 – A/Oth)

TO: Abstract (Deutsch)

TA: Abstract (tedesco)

PE: Abstract (italiano)

Esempio n. 119 (ProgFT_Testo3 – A/Oth)

TO: [...] wie sie Balasubramanian (2021) beschrieben hat.

TAFT: [...] come ha descritto Balasubramanian (2021).

PEFT: [...] come descritto da Balasubramanian (2021).

Nell'esempio n. 119 si riscontra un errore ortografico nel trasferimento di un nome proprio, probabilmente dovuto a un problema di riconoscimento del *token* o di allineamento tra testo di partenza e testo di arrivo.

Registra un numero limitato di occorrenze anche la categoria A/Mech/Capit, che riguarda gli errori di trasferimento delle lettere maiuscole e presenta un solo errore nel testo del sistema *fine-tuned*; mentre nel testo tradotto da DeepL non si registrano casi.

Complessivamente, le differenze più rilevanti riguardano le categorie F/Coh/InPu (uso incoerente della punteggiatura) e A/MisTra/SemUn (traduzione semanticamente errata), per le quali il sistema *fine-tuned* registra un numero significativamente maggiore di occorrenze. Al contempo, il modello mostra prestazioni migliori in riferimento ad alcuni aspetti caratterizzanti del linguaggio scientifico, quali la realizzazione delle nominalizzazioni (F/SLP/Nomin) e la strutturazione dei riferimenti bibliografici (F/Coh/Cit). Si riscontra inoltre una gestione più accurata della terminologia, in particolare dei termini che non richiedono traduzione (A/Termi/DNT), e un minor ricorso all'inglese come lingua *pivot* (A/ForLa). Per altri aspetti caratterizzanti il linguaggio scientifico, il modello *fine-tuned* registra le stesse occorrenze del sistema generico, come nell'uso dell'equivalente corretto di termini (A/Termi/SemUn) e nella traduzione delle *Fügungen* (F/Lex/Fue). Rimane invece più problematica la categoria della scelta lessicale non adeguata al genere e all'ambito specialistico (F/Lex/LexCho), il cui numero di occorrenze è leggermente più elevato rispetto a DeepL, pur costituendo un aspetto critico per entrambi i sistemi. Inoltre, per il modello *fine-tuned* emergono ulteriori categorie problematiche: alcune presentano lo stesso numero di occorrenze di DeepL (ad esempio A/MisExpl per i casi di inadeguatezza dei contenuti per il destinatario del testo di arrivo e F/De/CharForm per il mancato rispetto delle convenzioni tipografiche), mentre altre registrano valori più elevati (ad esempio A/MisTra/WSD per la traduzione errata di parole polisemiche, G/Orth/Punct per gli errori di punteggiatura e F/Coh/CoRef per la mancata corrispondenza tra referenti citati in frasi diverse). Nonostante queste difficoltà, i dati indicano che il modello *fine-tuned* rappresenta potenzialmente una base promettente per ulteriori sviluppi finalizzati al miglioramento della qualità delle traduzioni scientifiche dal tedesco all'italiano nel contesto linguistico.

8.3.2 Analisi dell'*effort* temporale e tecnico

In riferimento al testo tradotto dal sistema *fine-tuned*, la fase di *post-editing* si è articolata in sette sessioni, con durate individuali di 01:45:10, 01:53:02, 02:44:17, 01:02:24, 03:05:50, 01:32:40 e 02:32:05, per un totale complessivo di 14:35:28, mentre il totale delle sessioni di *post-editing* per il testo generato da DeepL è 12:23:17. Entrambi i testi presentano la stessa estensione (6.120 *tokens*),

ma i tempi di correzione differiscono sensibilmente, evidenziando come le modifiche apportate al testo del modello adattato abbiano richiesto un tempo maggiore. Pertanto, prendendo in esame il numero di *tokens* e la durata del *post-editing*, si può dedurre che la sola estensione del testo non sia un indicatore sufficiente per stimare il *post-editing effort*.

Confrontando il numero di errori annotati e la durata del *post-editing*, si trova conferma di tale deduzione. Il testo prodotto dal sistema *fine-tuned* presenta un numero maggiore di errori rispetto a quello generato da DeepL (685 contro 550), dato coerente con i tempi di *post-editing* più elevati. Tuttavia, non è la quantità assoluta di errori a determinare l'*effort*, bensì la loro natura e complessità, che incidono direttamente sulla tipologia e sull'intensità degli interventi correttivi richiesti.

Infine, dal rapporto tra *tokens* e tempo di correzione emerge una produttività media di circa 420 *tokens*/ora per il sistema *fine-tuned* e di circa 495 *tokens*/ora per DeepL. Tali dati suggeriscono che l'output di DeepL è complessivamente più agevole da post-editare, mentre quello del modello adattato richiede un *effort* superiore, riconducibile alla qualità più variabile della traduzione prodotta.

In sintesi, emerge che il *post-editing effort* dipende principalmente dalla qualità dell'output della TA, piuttosto che dalla sua estensione o dal mero numero di errori.

Dall'analisi del report di Post-Edit Compare è emerso che il testo tradotto dal sistema *fine-tuned* si compone di 207 segmenti e 4996 parole. La quasi totalità, pari a 198 segmenti, è stata generata dal sistema di TA adattato, mentre solo 9 segmenti sono stati inseriti manualmente in assenza di una proposta della TA. Diversamente, il testo tradotto da DeepL presenta 223 segmenti, pur mantenendo lo stesso totale di parole (4.996). Come già affermato precedentemente, la differenza nel numero di segmenti può essere ricondotta a una diversa segmentazione operata dai due sistemi di TA. Anche nel caso di DeepL, la maggior parte del testo, 219 segmenti, è stata generata automaticamente, mentre 4 segmenti risultano inseriti manualmente. In entrambi i casi non emergono proposte dalla memoria di traduzione, dato che conferma che l'intero materiale oggetto di analisi proviene esclusivamente dall'output dei sistemi di TA. Tale assenza è riconducibile al funzionamento di Post-Edit Compare: lo strumento opera, infatti, a livello di progetto di traduzione e utilizza una versione iniziale unica, salvata prima dell'avvio del *post-editing*, quando la memoria era ancora vuota. Di conseguenza, l'analisi si basa esclusivamente su tale versione e non tiene conto dei *matches* eventualmente generati durante le successive fasi di correzione.

La Tabella 18 riporta la distribuzione dei segmenti e delle parole nelle diverse fasce di PEMp. Ciascuna colonna (segmenti, parole e percentuali) mette a confronto i risultati delle traduzioni del modello *fine-tuned* con quelle di DeepL.

Fasce di PEMp	Segmenti		Parole		Percentuali	
Sistema di TA	Modello fine-tuned	DeepL	Modello fine-tuned	DeepL	Modello fine-tuned	DeepL
100%	5	39	100	592	2,08%	11,85%
95-99%	29	39	770	1.066	15,99%	21,34%
85-94%	75	44	1996	1.480	41,44%	26,42%
75-84%	32	18	797	510	16,55%	10,21%
50-74%	26	26	749	675	15,55%	13,51%
New (0-49%)	25	57	405	833	8,41%	16,67%

Tabella 16: Distribuzione comparativa delle fasce di PEMp per testo del modello *fine-tuned* n. 3

L'analisi comparativa delle fasce di PEMp evidenzia differenze significative nella qualità dell'output, misurata in termini di entità delle modifiche richieste durante il *post-editing*. Per quanto riguarda la fascia 100%, il modello *fine-tuned* presenta un numero estremamente contenuto di segmenti riutilizzati integralmente (5 segmenti, pari al 2,08% delle parole), mentre DeepL raggiunge un valore nettamente superiore (39 segmenti, 11,85% delle parole). Questo dato suggerisce che solo una parte estremamente marginale dell'output del sistema adattato risulta immediatamente accettabile, mentre DeepL riesce a produrre porzioni di testo più frequentemente conformi alla qualità richiesta.

Un andamento simile emerge nella fascia 95-99%, che raccoglie i segmenti sottoposti a correzioni minime. Anche in questo caso DeepL mostra valori più elevati (39 segmenti, 21,34% delle parole) rispetto al modello *fine-tuned* (29 segmenti, 15,99% delle parole), ciò segnala nel testo di DeepL segmenti di maggiore qualità che richiedono interventi estremamente contenuti.

La fascia 85-94% costituisce la porzione più consistente in entrambi gli output, ma con differenze rilevanti: il modello *fine-tuned* concentra in questa categoria il 41,44% delle parole (75 segmenti), mentre DeepL si attesta al 26,42% (44 segmenti). Ciò suggerisce che il sistema adattato genera una quota maggiore di output recuperabile con interventi relativamente lievi, ma richiede un *post-editing* costante; mentre DeepL distribuisce più uniformemente gli interventi necessari.

Le fasce intermedie, 75-84% e 50-74%, mostrano una distribuzione relativamente stabile nel modello *fine-tuned* (rispettivamente 16,55% e 15,55%), mentre DeepL presenta valori inferiori (10,21% e 13,51%), confermando che l'output del modello adattato necessita più frequentemente interventi di entità medio-alta.

La fascia *New* (0-49%), indicativa di segmenti completamente riscritti o con modifiche quasi integrali, rappresenta un indicatore critico della qualità dell'output. Il modello *fine-tuned* registra un valore dell'8,41% delle parole (25 segmenti), mentre DeepL raggiunge il 16,67% (57 segmenti). Sebbene non rappresenti la percentuale più elevata, questo dato conferma la tendenza emersa nelle

precedenti analisi: DeepL, pur producendo complessivamente più segmenti ad alta qualità, riscontra problematiche più critiche in una parte circoscritta dei casi, generando proposte di traduzione del tutto inaffidabile e soluzioni inadeguate.

Nel complesso, la distribuzione delle fasce di PEMP mette in luce due dinamiche distinte: DeepL fornisce una maggiore quantità di segmenti pronti all'uso o facilmente correggibili (100% e 95-99%), riducendo il *post-editing effort* nelle fasce superiori; tuttavia, quando non propone segmenti di alta qualità, è più probabile che generi segmenti che richiedono interventi di riscrittura estremamente consistenti. Invece, il modello *fine-tuned* concentra gran parte dell'output nella fascia 85-94%, richiedendo interventi sistematici e costanti ma meno estremi. L'eterogeneità complessiva di entrambe le distribuzioni conferma che, pur offrendo un supporto utile, i sistemi di TA non eliminano la necessità di un *post-editor* umano competente: più del 75% dell'output complessivo di entrambi i sistemi richiede comunque interventi più o meno rilevanti.

8.4 Riflessione conclusiva

L'analisi comparativa dei tre testi evidenzia alcune tendenze comuni riguardo alla qualità dell'output dei due sistemi di TA, nonché differenze specifiche legate alle singole caratteristiche testuali. In tutti i casi, il modello *fine-tuned* mostra un numero sensibilmente più elevato di occorrenze in alcune categorie problematiche ricorrenti, in particolare quelle legate alla coerenza testuale e alla correttezza semantica (ad esempio F/Coh/InPu e A/MisTra/SemUn) o alla scorrevolezza del testo (F/Sty/Disfl). Parallelamente, emergono però anche punti di forza del modello *fine-tuned*, comuni ai tre testi. In particolare, il sistema risulta più accurato in riferimento alla realizzazione delle nominalizzazioni (F/SLP/Nomin), alla strutturazione dei riferimenti bibliografici (F/Coh/Cit) e all'uso delle forme impersonali (F/SLP/ForImp), nonché nella gestione di termini che non richiedono traduzione (A/Termi/DNT) e nel minor ricorso all'inglese come lingua *pivot* (A/ForLa). Questi risultati indicano che il modello adattato può offrire un supporto solido nella resa di elementi strutturali e terminologici del linguaggio scientifico. Tuttavia, si osservano categorie particolarmente problematiche che registrano un numero maggiore di occorrenze rispetto al sistema generico, come illustrato nell'analisi degli errori dei singoli testi. È possibile individuare anche categorie per le quali i due sistemi registrano dei risultati analoghi, evidenziando, a seconda del testo analizzato, problematiche ricorrenti per entrambi. Questi dati confermano quanto osservato nelle precedenti analisi: la TA da sola non garantisce uniformità né adeguatezza nella traduzione dei testi scientifici. L'attività del *post-editor* rimane fondamentale per correggere errori di varia natura. Allo stesso tempo, i punti di forza del modello *fine-tuned* in alcuni aspetti strutturali e terminologici indicano che un approccio ibrido, che integri interventi umani con

sistemi adattati, costituisce potenzialmente una soluzione efficace per ottenere traduzioni affidabili e di qualità elevata.

Nel confronto tra i tre testi analizzati, emergono differenze significative sia in termini di *effort* temporale sia in relazione alla qualità dell'output generato dai sistemi di TA. Il tempo totale di *post-editing* relativo al testo di *Interculture journal* ammonta a 14:35:28 per il testo prodotto dal modello *fine-tuned* e a 12:23:17 per quello generato da DeepL; per il testo di *trans-kom* le durate complessive sono pari a 15:56:36 per il modello *fine-tuned* e 10:57:55 per DeepL; infine, per il contributo di *Informationen Deutsch als Fremdsprache* il *post-editing* ha richiesto 15:33:38 per il sistema adattato e 13:34:09 per DeepL. Tra i tre testi, *trans-kom* risulta quello più dispendioso in termini di tempo complessivo per il modello *fine-tuned*, mentre *Interculture journal* presenta la differenza più contenuta tra i due sistemi. In tutti i casi, tuttavia, l'output del modello adattato ha richiesto tempi di correzione più elevati.

Anche il numero di errori annotati fornisce un'indicazione utile per interpretare le differenze osservate. Il testo di *Interculture journal* presenta 685 errori nel testo del modello *fine-tuned* contro 550 in quello di DeepL; il testo di *trans-kom* mostra rispettivamente 1.141 e 794 errori; quello di *Informationen Deutsch als Fremdsprache* registra 1.048 errori per il sistema adattato e 800 per DeepL. La tendenza risulta coerente con quanto rilevato nelle durate del *post-editing*: in tutti i casi, il testo generato dal modello *fine-tuned* richiede un numero maggiore di interventi rispetto all'output di DeepL. Tuttavia, la quantità di errori non sempre si traduce in tempi di correzione proporzionali. Ad esempio, il testo di *Interculture journal* presenta meno errori complessivi rispetto agli altri due testi, tuttavia, si riscontra una differenza temporale relativamente ridotta tra i due sistemi, suggerendo che la complessità degli errori incida maggiormente della loro frequenza.

In riferimento alla lunghezza dei testi, emerge una distribuzione differenziata: il testo di *trans-kom* è quello più esteso con 8.484 *tokens*, seguito dal testo di *Informationen Deutsch als Fremdsprache* (8.454 *tokens*) e, infine, quello di *Interculture journal* (6.120 *tokens*). Nonostante ciò, l'estensione del testo non appare direttamente correlata al tempo richiesto per il *post-editing*. Il caso del testo di *Informationen Deutsch als Fremdsprache* è particolarmente indicativo: sebbene presenti una lunghezza inferiore rispetto al testo di *trans-kom*, ha richiesto un tempo di correzione superiore. Analogamente, il testo di *Interculture journal*, pur essendo il testo meno esteso, presenta un *effort* temporale non proporzionalmente ridotto. Ciò conferma che l'estensione del testo rappresenta un parametro informativo utile, ma non sufficiente, da solo, a spiegare l'*effort* effettivo.

Nel complesso, dall'analisi dei tre testi emerge che la qualità dell'output della TA costituisce il principale determinante del *post-editing effort*. Le differenze osservate tra i due sistemi, così come tra i tre testi, mostrano che né la lunghezza del testo né il semplice conteggio

degli errori risultano pienamente predittivi del carico di correzione richiesto. Piuttosto, la natura, la gravità e la distribuzione interna delle problematiche traduttive influenzano in modo decisivo i tempi di intervento.

L'analisi comparativa delle fasce di PEMP relative ai tre testi evidenzia pattern distinti e dinamiche operative che influenzano in modo significativo il *technical effort*. I dati confermano come la distribuzione delle parole del testo nelle diverse fasce non sia uniforme, mostrando differenze rilevanti tra i sistemi di TA, ma anche analogie nelle tendenze generali dei tre testi.

Per quanto riguarda la fascia 100%, che indica segmenti immediatamente riutilizzabili senza alcuna modifica, DeepL si distingue nettamente in tutti e tre i testi: per *Interculture journal* raggiunge l'11,85%, per *trans-kom* il 27,21%, e per *Informationen Deutsch als Fremdsprache* il 16,70%. Al contrario, il modello *fine-tuned* presenta valori molto più contenuti (rispettivamente 2,08%, 6,27% e 4,67%), confermando che solo una parte marginale dell'output prodotto dal sistema adattato è già utilizzabile senza interventi. La fascia 95-99%, dedicata a correzioni minime, mostra invece, una situazione più equilibrata: DeepL mantiene valori elevati nel caso di *Interculture journal* (21,34%) e *Informationen Deutsch als Fremdsprache* (13,39%), mentre nel testo di *trans-kom* il modello *fine-tuned* presenta un valore leggermente più alto di quello di DeepL (7,37% contro 6,39%). Nel complesso, il modello *fine-tuned* ottiene comunque percentuali più contenute nelle altre due riviste (15,99% per *Interculture journal* e 12,27% per *Informationen Deutsch als Fremdsprache*), suggerendo una minore incidenza di segmenti che richiedono interventi minimi rispetto a DeepL.

Successivamente, la fascia 85-94% rappresenta la porzione più consistente dell'output del sistema adattato in tutti e tre i testi: 41,44% per *Interculture journal*, 21,32% per *trans-kom* e 29,41% per *Informationen Deutsch als Fremdsprache*. Ciò indica che il modello *fine-tuned* tende a concentrare la maggior parte delle parole in una zona intermedia che richiede interventi lievi ma non trascurabili, mentre DeepL distribuisce le parole in maniera più bilanciata, riducendo la necessità di correzioni ricorrenti in questa fascia intermedia. Le fasce 75-84% e 50-74% confermano ulteriormente questa tendenza: il modello *fine-tuned* mostra valori generalmente più elevati rispetto a DeepL, con una concentrazione più consistente di parole che necessitano di interventi di entità medio-alta, evidenziando quindi una qualità più irregolare e meno prevedibile. Al contrario, DeepL mantiene una percentuale inferiore di parole in queste fasce, riducendo la necessità di correzioni di media portata. Infine, la fascia *New* (0-49%), che raccoglie segmenti integralmente riscritti o con modifiche ampiamente consistenti, mostra un comportamento peculiare. DeepL registra valori più elevati rispetto al modello *fine-tuned* in tutti i testi: 16,67% in *Interculture journal*, 35,87% in *trans-kom* e 22,32% in *Informationen Deutsch als Fremdsprache*.

Ciò indica che, sebbene DeepL produca segmenti di qualità superiore come indicato dalle fasce più elevate, nei casi problematici può produrre risultati particolarmente inadeguati, che richiedono una riscrittura completa o quasi totale. Tale tendenza emerge chiaramente anche dall'analisi qualitativa degli errori: gli errori di A/Omis, ad esempio, riguardano spesso segmenti ampi o frasi complete (cfr. esempio n. 96 in 8.3.1), mentre i casi di A/Add mostrano come DeepL tenda a introdurre aggiunte improprie particolarmente estese e strutturalmente invasive (cfr. esempio n. 72 in 8.2.1 ed esempio n. 106 in 8.3.1). Il modello *fine-tuned*, invece, si caratterizza, da un lato, per una netta prevalenza di segmenti collocati nella fascia intermedia e, dall'altro, per un numero più contenuto di casi appartenenti alla fascia *New* (0-49%). Ne risulta una qualità complessivamente più uniforme, ma mediamente inferiore, dovuta alla minore presenza di segmenti di alta qualità.

Nel complesso, le traduzioni dei tre testi mostrano due dinamiche complementari: DeepL tende a polarizzare l'output tra segmenti di elevata qualità e segmenti completamente inadeguati, mentre il modello *fine-tuned* mantiene una distribuzione più centrata sulle fasce intermedie, con interventi sistematici ma meno estremi. In tutti e tre i casi e in entrambi i sistemi, tuttavia, più del 75% dell'output richiede correzioni, confermando che il *post-editing* umano rimane indispensabile. La comparazione mette in evidenza come la scelta del sistema di TA influenzi non solo la qualità dell'output, ma anche la distribuzione degli interventi necessari, con impatti diretti sull'efficienza e sull'*effort* complessivo del *post-editor*.

Alla luce delle riflessioni emerse, è opportuno sottolineare che molte delle considerazioni già formulate nel capitolo 6 sul *post-editing effort* si riconfermano anche in questo contesto. In particolare, sebbene i dati analizzati forniscano informazioni utili sul *post-editing effort*, essi non coprono tutte le dimensioni: il *cognitive effort*, cruciale per comprendere il carico cognitivo e intellettuale del *post-editor*, non è stato valutato; mentre per quanto riguarda il *technical effort*, i report di Post-Edit Compare misurano solo le differenze tra la versione iniziale dell'output della TA e quella finale post-editata. In questo modo non vengono registrate eventuali modifiche intermedie, correzioni provvisorie o riformulazioni multiple sullo stesso segmento, necessarie per raggiungere la versione finale, tralasciando quindi una parte significativa dell'*effort* reale. Inoltre, permane la rilevanza della soggettività individuale: le scelte del *post-editor*, l'esperienza, la familiarità con il dominio e con il genere testuale, così come la propensione a intervenire più o meno intensamente, fattori che possono influenzare significativamente il tempo e la qualità della revisione.

Analogamente a quanto osservato nel capitolo 6, anche in questa analisi risulta evidente come il supporto umano sia indispensabile per garantire accuratezza, coerenza terminologica e leggibilità dei testi. L'integrazione di strumenti CAT, come Trados Studio con plug-in quali Post-Edit Compare e Quality, e di risorse esterne, come dizionari, banche dati di testi paralleli e

strumenti di analisi linguistica, continua a rappresentare un elemento fondamentale per ottimizzare il processo di *post-editing*, ridurre l'*effort* e assicurare un output di qualità pubblicabile. Questi risultati confermano, ancora una volta, che l'affidamento esclusivo alla traduzione automatica non è sufficiente, e che la combinazione tra intervento umano altamente qualificato e strumenti traduttivi tecnologici rimane centrale per garantire standard elevati per la traduzione di testi scientifici.

In conclusione, va sottolineato che i risultati ottenuti con il modello *fine-tuned* appaiono particolarmente incoraggianti se si considera la limitata quantità di dati impiegata per l'addestramento. Ciò lascia presumere che un ampliamento e una maggiore diversificazione del *training dataset* potrebbero generare miglioramenti sostanziali nelle prestazioni, con una conseguente riduzione del *post-editing effort* e un incremento della quota di segmenti di alta qualità. Tuttavia, è importante evidenziare che la qualità complessiva dell'output del modello *fine-tuned* non risulta costantemente superiore a quella di DeepL se si analizzano le singole categorie di errore; per alcune categorie specifiche, infatti, il sistema adattato registra un numero più elevato di criticità. Tale risultato deve comunque essere interpretato alla luce della differente disponibilità di risorse: DeepL è un sistema commerciale addestrato su quantità di dati ampie e non pubblicamente note, mentre il modello *fine-tuned* qui esaminato opera su un corpus significativamente più ridotto. Alla luce di queste considerazioni, le prestazioni osservate possono essere interpretate come una base promettente per sviluppi futuri, indicando che interventi mirati di ottimizzazione e l'espansione del corpus di addestramento potrebbero portare a un miglioramento significativo della qualità della TA del linguaggio scientifico tedesco in italiano in ambito linguistico.

Conclusioni e prospettive

In seguito all'ampia diffusione delle tecnologie per la traduzione, l'attività traduttiva si è progressivamente focalizzata sull'interazione tra l'essere umano e la macchina. Ciò ha contribuito all'affermazione della figura del *post-editor* come professionista sempre più richiesto nel mondo del lavoro contemporaneo (cfr. ELIS 2025).

La domanda di ricerca alla base del presente elaborato ha riguardato la valutazione della qualità della traduzione automatica dal tedesco all'italiano di contributi scientifici in ambito linguistico, nonché la determinazione del grado di intervento necessario da parte del *post-editor* sull'output della TA per raggiungere una traduzione di qualità pubblicabile.

A tal fine, sono stati condotti due studi sulla traduzione automatica del linguaggio scientifico tedesco in italiano: uno studio empirico basato sull'annotazione degli errori e sul *post-editing* di output generati da DeepL, da cui è stata elaborata la tassonomia TETASTI, e uno studio sperimentale volto a valutare l'impatto del *fine-tuning* di un sistema di TA sulla qualità della traduzione. I risultati di entrambi gli studi sono discussi nel paragrafo seguente.

9.1 Risultati della ricerca

In riferimento allo studio empirico riguardante la traduzione del linguaggio scientifico tedesco in italiano con DeepL, il primo risultato è costituito dalla stesura delle linee guida TETASTI che sono state sviluppate con l'obiettivo di creare uno strumento mirato per la coppia linguistica tedesco-italiano, il genere testuale dell'articolo scientifico e il dominio disciplinare della linguistica. Per la loro definizione sono stati assunti come riferimento i seguenti modelli consolidati di valutazione della qualità traduttiva: *Korrekturrichtlinien dell'Istituto per Traduttori e Interpreti dell'Università di Heidelberg* (IÜD, Universität Heidelberg 1996), *Annotation Guidelines for English-Dutch Translation Quality Assessment, version 1.1* (Daems et al. 2013), *Annotation Guidelines for legal Italian<->(South Tyrolean) German Machine Translation Quality Assessment* (Chiocchetti e De Camillis 2022), *SCATE taxonomy* (Tezcan et al. 2017) e *Multidimensional Quality Metrics (MQM)* (Lommel, Uszkoreit, et al. 2014; Lommel et al. 2013). La struttura di TETASTI riprende la distinzione tra *Accuracy* e *Fluency* presente nella *SCATE taxonomy* e nelle *Annotation Guidelines* di Chiocchetti e De Camillis (2022), integrandola con la dimensione di *Grammar* introdotta dalle *Korrekturrichtlinien dell'Istituto per Traduttori e Interpreti dell'Università di Heidelberg*. Gli errori di *Accuracy* riguardano la corrispondenza informativa tra testo di partenza e testo di arrivo; quelli di *Fluency* concernono la scorrevolezza e la naturalezza della lingua di arrivo; mentre quelli di

Grammar ne riguardano la correttezza formale. Tuttavia, rispetto ai modelli esistenti, TETASTI presenta un adattamento originale utile a descrivere in modo più adeguato fenomeni specifici del linguaggio scientifico. In particolare, sono state introdotte nuove categorie, tra le quali si individuano:

- Categorie per i fenomeni del linguaggio scientifico: F/SLP/Nom per la mancata realizzazione di nominalizzazioni e F/SLP/ForImp per l'impiego errato di forme personali al posto delle forme impersonali;
- Categoria di F/Lex/Fue per gli errori nella traduzione delle *Fügungen*, combinazioni idiomatiche tipiche del linguaggio scientifico;
- Categoria di F/Coh/Cit per segnalare la strutturazione incoerente dei riferimenti bibliografici;
- Categorie per gli aspetti di layout e tipografici: A/Mech/CharFor per gli errori di trasferimento della formattazione; F/De/CharFor per la formattazione non coerente con le convenzioni tipografiche della lingua di arrivo; e F/De/Lay per errori nella strutturazione del testo;
- Categoria di F/Coh/InPu per segnalare l'uso incoerente della punteggiatura.

Sono state inoltre riadattate alcune categorie, tratte da linee guida preesistenti, le cui definizioni sono state modificate al fine di descrivere con maggiore precisione le caratteristiche del genere testuale, dell'ambito specialistico e della coppia linguistica considerati, nonché gli errori emersi nel corpus analizzato. Tra queste rientrano:

- Categorie per la componente terminologica: A/Termi/SemUn per la traduzione errata di termini e A/Termi/DNT per la traduzione non necessaria di termini;
- Categoria di F/Lex/LexCho errori riguardanti la scelta lessicale maggiormente adeguata all'ambito specialistico e al genere testuale preso in esame;
- Categorie riguardanti i lessemi composti: A/MWE/Komp per la traduzione errata di aggettivi composti e A/MWE/Ell per gli errori nella traduzione di più parole composte caratterizzate dall'elissi di un elemento che si ripete.

La stesura della tassonomia TETASTI ha evidenziato i limiti delle linee guida di categorizzazione e annotazione degli errori generali, che non risultano pienamente adeguate a descrivere i fenomeni propri del linguaggio e del genere testuale analizzato. L'adozione di linee guida che rispondono a tali necessità ha l'intento di facilitare il lavoro del *post-editor* nella valutazione qualitativa dell'output della TA, consentendogli di determinare con maggiore precisione e consapevolezza la pertinenza dei propri interventi, distinguendo tra le correzioni necessarie a

soddisfare i requisiti di qualità e quelle che, invece, eccederebbero i confini propri dell'attività del *post-editor*.

Nello studio empirico presentato nei capitoli 5 e 6, la qualità della traduzione generata da DeepL è stata valutata sulla base delle annotazioni di errori, del tempo impiegato per lo svolgimento del *post-editing* e della quantità di modifiche apportate ai segmenti del testo. L'analisi degli errori ha evidenziato difficoltà da parte del sistema di TA considerato nella gestione terminologica, con particolare riferimento alla selezione dell'equivalente corretto nella lingua di arrivo (A/Termi/SemUn) o al mancato riconoscimento dell'uso consolidato di termini in lingua inglese (A/Termi/DNT). Altre criticità hanno riguardato la strutturazione dei riferimenti bibliografici (F/Coh/Cit), la realizzazione delle nominalizzazioni (F/SLP/Nom), la scelta lessicale adeguata al dominio trattato (F/Lex/LexCho), l'impiego delle forme impersonali (F/SLP/ForImp) o di combinazioni consolidate nell'ambito del linguaggio scientifico (*Fügungen*, F/Lex/Fue). Inoltre, è stata riscontrata la tendenza di DeepL a tradurre ogni segmento testuale, senza distinguere tra contenuto comunicativo e contenuto metalinguistico. In tali casi, il *post-editor* deve riconoscere che la frase non appartiene al discorso dell'autore, ma costituisce un elemento da mantenere nella lingua utilizzata nel testo di partenza, al fine di preservare l'accuratezza e la validità scientifica del testo (cfr. ad es. Capitolo 6 - esempio n. 14). Analogamente, il sistema ha mostrato dei limiti nella gestione delle citazioni in tedesco per le quali esistono traduzioni ufficiali in italiano, ignorate dal modello (cfr. ad es. Capitolo 6 - esempio n. 57). Resta infine fondamentale sottolineare che anche gli errori relativi alle altre categorie, sebbene per definizione non afferiscano direttamente al linguaggio scientifico, possono influenzare significativamente la qualità del testo e la resa del linguaggio scientifico nei suoi elementi costitutivi.

Per quanto riguarda il tempo impiegato per lo svolgimento del *post-editing*, anche definito come *temporal effort*, i dati sono stati ottenuti mediante l'uso del plug-in di Trados Studio Quality. L'analisi comparativa tra le tempistiche di *post-editing*, la quantità di errori annotati e le estensioni dei contributi ha rivelato una correlazione non lineare tra queste variabili: contrariamente alle aspettative, la dimensione testuale e il numero di errori annotati non sono sempre stati direttamente proporzionali al tempo di *post-editing* richiesto. Si è osservato, a titolo esemplificativo, come i progetti di traduzione *Informationen Deutsch als Fremdsprache* e *Interculture journal*, pur essendo di minore estensione, abbiano richiesto un tempo di *post-editing* superiore al progetto *trans-kom*, che presenta testi di lunghezza maggiore. Tale divergenza ha trovato conferma anche nel rapporto tra annotazioni e durata del processo: il rapporto tra errori annotati e tempo impiegato è risultato proporzionale per il progetto *Interculture journal*, mentre si è invertito per *Informationen Deutsch als Fremdsprache* e *trans-kom*, dove quest'ultimo ha presentato più annotazioni nonostante

la durata di *post-editing* inferiore. Nel complesso, questi dati hanno suggerito che il *temporal effort* dipende principalmente dalla qualità dell'output della TA e dalla tipologia degli errori, piuttosto che dalla sola estensione del testo o dal numero complessivo di errori.

Il *technical effort* è stato quantificato attraverso il parametro della *Post-Edit Modifications percent* (PEMp), ottenuta mediante il plug-in di Trados Studio Post-Edit Compare, che indica la percentuale di parole riutilizzate senza modifiche. L'analisi di tali dati ha delineato un'attività di *post-editing* per i tre progetti eterogenea e sbilanciata: accanto a segmenti pronti all'uso (fascia 100%), sono state necessarie correzioni di varia entità, fino a rielaborazioni integrali o quasi complete che hanno comportato un *effort* considerevole (fascia *New*, 0-49%) e i cui valori sono risultati paragonabili o superiori a quelli registrati per i segmenti con qualità già pubblicabile. In tutti i progetti, la fascia maggiormente rappresentata è stata 85-94%. Per le restanti fasce, i progetti hanno registrato percentuali variabili, rendendo difficile l'individuazione di una tendenza univoca.

Complessivamente, i risultati hanno confermato che, nonostante i progressi tecnologici, l'affidamento esclusivo alla TA non garantisce uniformità né adeguatezza nella traduzione dei contributi scientifici. Il *post-editing* rimane quindi indispensabile per gestire la forte variabilità qualitativa dell'output, per raggiungere una qualità pubblicabile e garantire coerenza e leggibilità complessiva.

Lo studio sperimentale, illustrato nei capitoli 7 e 8, ha trattato il *fine-tuning* di un sistema pre-addestrato mediante l'uso di un *syntethic corpus*, la cui componente sintetica corrisponde alla lingua di partenza ottenuta traducendo automaticamente in tedesco articoli scientifici italiani, che costituiscono invece la parte di arrivo del corpus. Mediante questo modello adattato sono stati tradotti in italiano tre articoli tedeschi già analizzati nello studio precedente e queste traduzioni sono state confrontate con quelle generate da DeepL.

Dall'analisi comparativa è emerso che, in tutti i casi, il modello *fine-tuned* ha presentato un numero maggiore di errori in alcune categorie ricorrenti, in particolare quelle legate all'uso incoerente della punteggiatura (F/Coh/InPu) e alla correttezza semantica della traduzione (A/MisTra/SemUn) o alla scorrevolezza della lingua di arrivo (F/Sty/Disfl). Parallelamente, il modello ha mostrato punti di forza condivisi nei tre testi: in particolare, rispetto al sistema di TA DeepL, è risultato più accurato nella resa delle nominalizzazioni (F/SLP/Nomin), ha gestito meglio la strutturazione dei riferimenti bibliografici (F/Coh/Cit), l'uso delle forme impersonali (F/SLP/ForImp), ha riconosciuto in maniera più sistematica i termini che non richiedono traduzione (A/Termi/DNT) e ha ridotto l'uso dell'inglese come lingua *pivot* (A/ForLa). Alcune categorie, tuttavia, sono risultate particolarmente problematiche per il modello *fine-tuned* e hanno presentato più errori rispetto al sistema generico, come evidenziato nell'analisi dei singoli testi. È stato

possibile individuare anche categorie per le quali i due sistemi hanno registrato risultati analoghi, mostrando, a seconda del testo analizzato, criticità persistenti per entrambi (cfr. Capitolo 8).

In merito all'applicazione delle linee guida di annotazione, la tassonomia TETASTI si è dimostrata uno strumento di analisi solido e adeguato alle specificità del corpus considerato. Ciò è stato confermato dal fatto che non si è resa necessaria alcuna integrazione o modifica delle categorie proposte, che sono risultate esaustive e operativamente chiare per la classificazione degli errori negli output analizzati.

Come già affermato, in tutti e tre i testi tradotti dal modello *fine-tuned*, il numero di annotazioni è stato superiore rispetto a quello dei testi generati da DeepL. Tale tendenza è risultata coerente con i dati relativi al *temporal effort*, secondo cui gli output del modello adattato hanno richiesto tempi di correzione più elevati. Tuttavia, dall'analisi è emerso che né la quantità di errori né l'estensione del testo sono stati direttamente proporzionali alla durata del *post-editing*: ciò che ha influenzato in modo decisivo il *post-editing effort* è stata di fatto la natura, la gravità e la distribuzione interna delle problematiche traduttive, ossia più in generale la qualità dell'output della TA.

Per quanto concerne l'analisi comparativa delle fasce di PEMP, è emerso che, sebbene DeepL abbia prodotto un numero più elevato di segmenti già pubblicabili (fascia 100%) o che hanno richiesto interventi minimi (95-99%), nei casi problematici ha prodotto risultati particolarmente inadeguati, con un'incidenza significativa di segmenti che hanno richiesto una riscrittura completa o quasi integrale (fascia *New*, 0-49%). Ciò è stato confermato anche dall'analisi qualitativa degli errori: gli errori di omissioni (A/Omis), ad esempio, hanno riguardato spesso segmenti ampi o frasi complete (cfr. Capitolo 8 - esempio n. 96), mentre i casi di aggiunte improprie (A/Add) hanno mostrato come DeepL abbia tendenzialmente introdotto aggiunte improprie particolarmente estese e strutturalmente invasive (cfr. Capitolo 8 - esempio n. 72 e 106). Il modello *fine-tuned* invece, pur concentrando gran parte degli interventi di *post-editing* nelle fasce intermedie (85-94% e 75-84%), ha registrato meno casi nella fascia *New*, offrendo una qualità più uniforme ma mediamente inferiore. In sintesi, DeepL tende a polarizzare la qualità producendo sia segmenti ottimali che segmenti totalmente inadeguati, mentre il modello *fine-tuned* mantiene una distribuzione degli interventi più centrata sulle fasce intermedie, con azioni sistematiche ma meno estreme. In tutti e tre i testi, tuttavia, più del 75% dell'output richiede correzioni.

I dati hanno confermato dunque quanto osservato nelle precedenti analisi: la TA non garantisce, autonomamente, l'uniformità e l'adeguatezza richiesta dai testi scientifici, rendendo indispensabile l'attività del *post-editor* per correggere errori di varia natura. Ciononostante, i punti di forza del modello *fine-tuned* in alcuni aspetti strutturali e terminologici suggeriscono che un

approccio ibrido, che combini interventi umani con sistemi adattati, costituisce potenzialmente una soluzione efficace per ottenere traduzioni affidabili e di qualità elevata. Nel complesso, i risultati mostrano che entrambi i sistemi presentano punti di forza e criticità, e che la qualità dell'output varia sensibilmente in funzione dei fenomeni linguistici coinvolti, così come il tempo richiesto per il *post-editing*, che si conferma elevato, evidenziando come l'impiego della TA non comporti necessariamente un risparmio di tempo rispetto alla traduzione interamente umana.

Per entrambi gli studi, l'impiego di Trados Studio come ambiente di lavoro si è rivelato fondamentale per una gestione più agevole del *workflow*. L'integrazione di plug-in, quali QualityTools e Post-Edit Compare, non solo ha fornito un supporto operativo, ma ha anche messo a disposizione dati quantitativi indispensabili per un'analisi oggettiva dell'interazione tra output della TA e intervento di *post-editing*. Inoltre, la possibilità di creare memorie di traduzione e database terminologici, unitamente alla funzione ricerca concordanze del CAT tool, ha facilitato il recupero di segmenti o singoli elementi già tradotti, accelerando il processo di traduzione e garantendo coesione e coerenza all'interno del testo. Parallelamente agli strumenti integrati nel CAT tool, si sono rivelate essenziali anche le risorse esterne (ad esempio Sketch Engine, dizionari monolingui e bilingui ecc.) consultate per svolgere il *post-editing* durante entrambi gli studi. Tali risorse hanno supportato il *post-editor* nella formulazione di alternative traduttive e nelle verifiche relative alla correttezza formale e all'accuratezza semantica delle proposte generate dal sistema di TA.

9.2 Risposta alle domande di ricerca

Le domande di ricerca al centro del presente elaborato sono le seguenti: qual è la qualità dell'output grezzo generato dal sistema di TA? E, di conseguenza, quanto lavoro è richiesto al *post-editor* per elaborare proposte di traduzione utili al raggiungimento di una qualità pubblicabile? Alla luce dei risultati discussi nel paragrafo precedente, è possibile fornire risposte articolate a entrambe le questioni.

In merito alla prima questione, i dati mostrano che la qualità dell'output della TA è complessivamente disomogenea e fortemente influenzata da specifici aspetti del linguaggio scientifico, come ad esempio la gestione della terminologia, la corretta resa di strutture linguistiche tipiche del discorso scientifico (come nominalizzazioni, forme impersonali o costrutti idiomatici tipici), nonché la coerenza interna, in particolare nelle citazioni e nei riferimenti bibliografici. A ciò si aggiunge la difficoltà riscontrata dal sistema nel fornire una traduzione che sia costantemente adeguata dal punto di vista del registro, della naturalezza e della correttezza grammaticale. Inoltre, sono stati riscontrati diversi casi problematici in cui il sistema di TA non distingue il contenuto comunicativo da quello metalinguistico (come esempi utili a esplicitare problematiche traduttive o

di apprendimento degli studenti), con conseguente compromissione dell'accuratezza e della fedeltà al testo di partenza. Ne deriva un output solo parzialmente affidabile poiché le proposte di traduzione formulate risultano estremamente variabili e richiedono in ogni caso l'intervento umano per verificarne la qualità e per apportare eventuali modifiche utili a garantire un risultato adeguato alla qualità attesa. Oltre agli aspetti prettamente linguistici e strutturali, emerge anche una criticità di ordine più generale: i sistemi di TA tendono a generare traduzioni apparentemente scorrevoli, che comportano un'"illusione di naturalezza" e possono celare deviazioni semantiche, inesattezze terminologiche o variazioni di registro. Per questo motivo il *post-editor* è chiamato a un controllo attento e analitico della correttezza semantica, della scelta lessicale e della coerenza concettuale. In particolare, nel linguaggio scientifico, in cui ogni lessema assume un valore determinante, una valutazione rigorosa di ciascuna parola è fondamentale per garantire la fedeltà al testo di partenza e la precisione contenutistica del risultato finale.

La seconda domanda di ricerca, relativa alla quantità di lavoro necessaria al *post-editor* per raggiungere una qualità pubblicabile, trova risposta nei dati relativi al *temporal* e al *technical effort*. Pur mostrando una certa variabilità tra testi e sistemi di TA, i risultati confermano che la quantità di lavoro richiesta non dipende in modo lineare dall'estensione del testo né dal numero complessivo di errori annotati: il fattore determinante risiede, piuttosto, nella natura degli errori, la loro distribuzione interna e il loro impatto sul testo e sul linguaggio scientifico. Nei progetti di traduzione analizzati, oltre il 75% dei segmenti richiede modifiche, spesso di entità rilevante. Ciò conferma che il *post-editing* rimane un passaggio imprescindibile ma anche gravoso.

Inoltre, i dati relativi all'*effort* temporale e tecnico sollevano interrogativi sulla reale efficienza e convenienza del *full post-editing*. Sebbene tale approccio consenta di ottenere un testo con qualità comparabile a quella di una traduzione umana, i risultati dei progetti analizzati mostrano che il tempo e lo sforzo tecnico richiesti possono avvicinarsi a quelli necessari per una traduzione *ex novo*. Ne consegue che il vantaggio in termini di efficienza tipico della traduzione automatica può ridursi significativamente, e che la scelta tra *post-editing* e traduzione umana deve essere valutata caso per caso, tenendo conto del genere testuale, della coppia linguistica e del dominio specialistico.

In questo quadro, assume particolare rilievo la distinzione tra *light post-editing* e *full post-editing*. I dati analizzati mostrano che, nel caso di articoli scientifici di linguistica, il *light post-editing* non è sufficiente a garantire una qualità pubblicabile, poiché tende a trascurare interventi di natura stilistica e testuale. Tuttavia, nel linguaggio scientifico, lo stile, inteso ad esempio come adeguatezza del registro, uso corretto delle forme impersonali, resa accurata delle *Fügungen*, preferenza per costruzioni nominali rispetto a quelle verbali, rappresenta un elemento

imprescindibile per la qualità complessiva del testo. Il *full post-editing* si è pertanto rivelato un passaggio necessario, seppur particolarmente oneroso in termini di tempo, sforzo cognitivo e tecnico. Tale onerosità appare tuttavia giustificata dalla complessità del linguaggio scientifico e dalla necessità di garantire accuratezza, coerenza e affidabilità del testo finale.

Un ulteriore aspetto emerso riguarda la componente di soggettività che permea l'intero processo di valutazione della qualità e di *post-editing*. L'individuazione degli errori, la selezione delle soluzioni traduttive, la propensione a intervenire in modo più o meno invasivo e la dimensione temporale sono fattori che dipendono dalle competenze e dall'esperienza del singolo *post-editor*. Tale variabilità suggerisce che il *post-editing* non sia un'attività meccanica, bensì un processo cognitivo complesso, in cui l'interazione tra competenza linguistica, familiarità con il dominio specialistico e giudizio personale svolge un ruolo determinante. In quest'ottica, il *post-editing* si distingue nettamente sia dalla traduzione che dalla revisione tradizionale. Come sottolineato da Veale e Way (1997), esso riguarda la correzione di un testo generato da un sistema automatizzato, e non l'intervento su una traduzione umana: ciò comporta vincoli specifici, tra cui una minore libertà creativa e un'attenzione costante alla coerenza logica e terminologica del sistema. Diversi studi hanno evidenziato come il *post-editing* tenda a limitare la creatività traduttiva, orientando il lavoro verso operazioni correttive e decisioni localizzate piuttosto che verso una riformulazione più ampia del testo (cfr. ad es. Arenas/Toral 2020).

Accanto alle competenze linguistiche e specialistiche, emergono anche fattori ergonomici e individuali che influenzano in modo significativo il lavoro del *post-editor*. Tra questi figurano l'importanza di un ambiente di lavoro e strumenti adeguati, così come la necessità di pause regolari per gestire il carico cognitivo. Come ricorda emblematicamente Paul Berf (2019 cit. in Schindler 2021: 47), "Routine ist ein großer Feind des Übersetzers", evidenziando come la ripetitività possa compromettere la concentrazione e la qualità delle decisioni traduttive.

Infine, il *post-editor* deve costantemente mantenere un equilibrio tra aderenza all'input e giudizio professionale: un'eccessiva rigidità può portare a correzioni superflue, mentre un atteggiamento estremamente permissivo può ridurre la qualità finale della traduzione. Questa oscillazione potenziale conferma la natura complessa e cognitivamente impegnativa del processo di correzione.

In sintesi, il *post-editing* si configura come un'attività a ridotta componente creativa ma ad alta intensità cognitiva, la cui efficacia dipende da variabili molteplici: la qualità dell'output della TA, le competenze e strategie del traduttore, la sua motivazione e il contesto operativo. A questi elementi si aggiungono fattori legati ai testi esaminati, quali la coppia linguistica, il dominio

disciplinare e le caratteristiche del linguaggio scientifico, che influenzano in modo significativo la complessità dell'intervento richiesto.

Alla luce di quanto esposto, risulta chiaro che, per la traduzione in italiano di articoli scientifici tedeschi di linguistica, l'impiego esclusivo della traduzione automatica non garantisce una qualità pubblicabile. L'intervento umano rimane quindi fondamentale, e la TA va considerata come uno strumento da integrare in un flusso di lavoro ibrido, dove memorie di traduzione, glossari e modelli adattati al dominio contribuiscano a migliorare efficienza e accuratezza del testo finale.

9.3 Prospettive per ulteriori ricerche

Quanto emerso dall'analisi dei risultati offre spunti utili per sviluppi futuri nella ricerca e nelle applicazioni pratiche della traduzione automatica. Una possibile direzione consiste nell'ampliare l'indagine includendo anche la dimensione del *cognitive effort*, che completa il quadro del carico richiesto al *post-editor* affiancandosi alle misure di *temporal* e *technical effort* adottate nel presente elaborato. L'integrazione di dati cognitivi, raccolti ad esempio attraverso i metodi di *eye-tracking* o *think aloud protocols* (cfr. 3.6), permetterebbe di comprendere in modo più approfondito i processi decisionali coinvolti (come la lettura e la comprensione del testo, l'individuazione degli errori, la formulazione di proposte di traduzione alternative ecc.) e l'impatto cognitivo generato dalle diverse tipologie di errore.

Un ulteriore tema che potrebbe essere approfondito in studi futuri riguarda la questione, ancora largamente irrisolta, della retribuzione del *post-editor*. Poiché il *post-editing* richiede competenze linguistiche e specialistiche elevate e implica un *effort* articolato, rimane aperta la questione di come quantificarne in modo equo il valore economico. Una riflessione sistematica sui modelli di retribuzione, oggi non uniformi né pienamente definiti, costituirebbe un contributo rilevante sia per la ricerca sia per la pratica professionale, soprattutto alla luce dell'evoluzione delle tecnologie di TA e della crescente centralità del ruolo del *post-editor*.

In aggiunta, si potrebbe ampliare il corpus analizzato, includendo un numero maggiore di articoli scientifici tedeschi da tradurre appartenenti a differenti sottodomini della linguistica. Ciò consentirebbe non solo di verificare la generalizzabilità dei risultati qui presentati, ma anche di esplorare eventuali variazioni nella tipologia degli errori, nella loro distribuzione, nonché nel *post-editing effort*. L'ampliamento del corpus permetterebbe quindi di testare la robustezza e la capacità delle linee guida proposte di essere applicate in modo coerente e sistematico a corpora di dimensioni maggiori e di natura diversa, che, tenendo conto dei possibili sviluppi futuri della tecnologia, potrebbero presentare esigenze e problematiche traduttive differenti.

Un'analisi comparativa delle strategie di *post-editing* adottate da diversi *post-editor* potrebbe offrire una prospettiva più ampia sui comportamenti traduttivi e sulle preferenze operative. Esaminare come professionisti con diversa formazione linguistica, esperienza traduttiva o familiarità con il dominio intervengano sui medesimi segmenti, permetterebbe di delineare approcci più efficaci e sostenibili nella collaborazione essere umano-macchina, contribuendo alla definizione di possibili standard operativi o procedure di *post-editing* maggiormente sistematiche.

I risultati emersi dall'individuazione e dalla classificazione degli errori più ricorrenti costituiscono inoltre un punto di partenza significativo per lo sviluppo o il perfezionamento dei sistemi di TA. Le categorie di errore definite, in particolare quelle relative alla terminologia, alla coesione, alla resa delle strutture tipiche del linguaggio scientifico, offrono indicazioni concrete per interventi mirati di ottimizzazione. Questi risultati appaiono utili non solo per i *post-editor*, che possono così prevedere dove si concentreranno i loro interventi, ma anche per le grandi aziende che sviluppano sistemi di TA, in quanto forniscono indicazioni su dove la macchina tende a sbagliare, dove necessita di miglioramenti e dove invece produce output affidabili. In tal modo, è possibile indirizzare con maggiore efficacia le strategie di *training* o di adattamento dei modelli.

In tale prospettiva, dunque, il *fine-tuning* si conferma uno strumento promettente. Sebbene nello studio sperimentale qui riportato il modello adattato sia stato addestrato su un corpus di dimensioni limitate, i risultati mostrano tendenze incoraggianti per alcuni aspetti della traduzione del linguaggio scientifico. Oltre all'ampliamento del corpus oggetto di studio, si potrebbe considerare un'espansione e una maggiore diversificazione del *training dataset*: ciò potrebbe apportare miglioramenti sostanziali alla qualità complessiva dell'output, con una possibile riduzione del *post-editing effort* e un incremento della proporzione di segmenti direttamente utilizzabili.

È importante ricordare che la qualità complessiva dell'output del modello *fine-tuned* non risulta costantemente superiore a quella di DeepL. Tale risultato deve essere interpretato alla luce della differente disponibilità di risorse. Mentre i sistemi commerciali come DeepL sono addestrati su ampie quantità di dati, spesso ottenuti attraverso modalità di raccolta non pienamente note o documentate, i modelli *fine-tuned* derivanti dalla ricerca accademica devono rispettare vincoli più stringenti in termini di copyright, privacy e tracciabilità delle fonti e quindi dispongono di una quantità di dati significativamente più ridotta. Tale disparità nel volume e nella tipologia dei dati accessibili contribuisce a spiegare le differenze osservate nelle prestazioni dei due sistemi, costituendo un elemento interpretativo rilevante per la lettura comparativa dei risultati, ma al contempo solleva questioni etiche ancora aperte relative alla trasparenza dei processi di raccolta dei dati nei diversi contesti di sviluppo.

Infine, le strategie di adattamento per specifici domini specialistici rappresentano un'ulteriore direzione promettente. In questo senso, le memorie di traduzione e i glossari elaborati personalmente dal *post-editor* potrebbero rivelarsi utili come materiale per adattare i sistemi di TA in modo più mirato, consentendo ai *post-editor* di beneficiare di modelli ad uso personale più coerenti e accurati dal punto di vista lessicale e terminologico rispetto al dominio di interesse. Il potenziamento di modelli adattati specificamente al linguaggio scientifico tedesco-italiano potrebbe quindi rappresentare un'importante direzione di sviluppo, sia per la ricerca accademica sia per le applicazioni professionali.

Bibliografia

Letteratura primaria:

- Agnetta, Marco. 2021. «Kompensation und Redundanz. Zwei Prinzipien der Text-Bild-Beziehung in leicht verständlichen Kommunikaten». *trans-kom* 14 (2): 257–77.
- Barmeyer, Christoph I., e Eric Davoine. 2016. «Konstruktives interkulturelles Management – Von der Aushandlung zur Synergie». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 15 (26): 97–115. <https://doi.org/10.24403/jp.1278518>.
- Bolten, Jürgen. 2016. «Interkulturelle Trainings neu denken». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 15 (26): 75–91. <https://doi.org/10.24403/jp.1278516>.
- Brunsing, Theresa. 2016. «Landeskundliche Abbildungen in Lehrbüchern für Deutsch als Fremdsprache. Die Entwicklung des Bildeinsatzes in den Jahren 2000 bis 2010». *Informationen Deutsch als Fremdsprache* 43 (5): 494–515. <https://doi.org/doi:10.1515/infodaf-2016-0503>.
- Burger, Günter. 2019. «Was bewirkt der Einsatz von Spielfilmen im Fremdsprachenunterricht? Die empirische Forschung im Überblick». *Informationen Deutsch als Fremdsprache* 46 (6): 687–713. <https://doi.org/doi:10.1515/infodaf-2018-0078>.
- Funk, Hermann, Manja Gerlach, e Dorothea Spaniel. 2016. «Materialeinsatz in tandembasierten Fremdsprachenlernszenarien. Bericht aus einem laufenden Forschungsprojekt». *Informationen Deutsch als Fremdsprache* 43 (6): 693–709. <https://doi.org/doi:10.1515/infodaf-2016-0609>.
- Geist, Barbara. 2018. «Erhebung syntaktischer Fähigkeiten von Kindern mit Deutsch als Zweitsprache: Ein Methodenvergleich». *Informationen Deutsch als Fremdsprache* 45 (4): 423–43. <https://doi.org/doi:10.1515/infodaf-2018-0068>.
- Gruber, Valerie V. V., e Eberhard Rothfuß. 2016. «Interkulturelle Managementforschung – Reflexive Gedanken über eine unreflektierte Denkschule». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 15 (26): 117–40. <https://doi.org/10.24403/jp.1278519>.
- Henze, Jürgen. 2016. «Vom Verschwinden des (Inter) Kulturellen und Überleben der (Inter)Kulturalität». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 15 (26): 59–74. <https://doi.org/10.24403/jp.1278515>.
- Hilgert, Mariana. 2015. «Übersetzung zwischen möglichen Welten – eine Suche nach einem theoretischen Ausweg». *trans-kom* 8 (2): 539–59.
- Ivanova, Alina, Markus Kollmannsberger, e Ewald Kiel. 2017. «Herausfordernde Situationen im interkulturellen schulischen Kontext: Wahrnehmungs- und Handlungsmuster von Lehrkräften». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 16 (27–28): 23–41. <https://doi.org/10.24403/jp.1278337>.
- Kempa, Thomas. 2013. «Anregungen für eine Differenzierung eines funktionalen translationsdidaktischen Verstehens- und Sinnbegriffs aus pragmatistischer Sicht». *trans-kom* 6 (2): 345–70.

- Kriegel-Schmidt, Katharina, e Klaus Schmidt. 2014. «Kultur als Blickwinkel in interkulturell orientierten Lernsettings. Didaktische und theoretische Reflexionen am Beispiel einer Ausbildung in Interkultureller Mediation». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 13 (22): 27–44. <https://doi.org/10.24403/jp.1282887>.
- Longhi, Elisabetta. 2022. «Moorheilbäder für Italiener. Ein linguistischer Fall». *trans-kom* 15 (1): 98–127.
- Nazarkiewicz, Kirsten. 2013. «Interkulturalität als immanenter Faktor in Coaching und Training – konzeptionelle Überlegungen». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 12 (20): 47–67. <https://doi.org/10.24403/jp.1285138>.
- Nicklaus, Martina. 2020. «Wann klingt (übersetzte) Sprache fremd?» *trans-kom* 13 (2): 125–44.
- Niewalda, Katrin, Maria Gabriela Schmidt, e Shinichi Sakamoto. 2016. «Ergänzungsvorschläge für Hörverstehensübungen in deutschen Lehrwerken im universitären DaF-Unterricht in Japan». *Informationen Deutsch als Fremdsprache* 43 (6): 623–46. <https://doi.org/doi:10.1515/infodaf-2016-0606>.
- Pavlova, Anna. 2014. «Strategie der Übersetzung und Beurteilung der Übersetzungsqualität». *trans-kom* 7 (2): 256–71.
- Pust, Daniel. 2019. «Integrationsmechanismen des Blended Learning im Fremdsprachenunterricht. Eine explorative Fallstudie». *Informationen Deutsch als Fremdsprache* 46 (5): 586–601. <https://doi.org/doi:10.1515/infodaf-2018-0077>.
- Roncoroni, Tiziana. 2022. «Artikelgebrauch in Text und Übersetzung. Beobachtungen zum Sprachenpaar Deutsch-Italienisch». *trans-kom* 15 (1): 242–69.
- Runte, Maren. 2015. «Wie benutzen fortgeschrittene DaF-Lernende Wörterbücher? Eine Eye-Tracking-Studie zur Benutzung von Lernerwörterbüchern und ein Vorschlag zu deren Verbesserung». *Informationen Deutsch als Fremdsprache* 42 (5): 476–98. <https://doi.org/doi:10.1515/infodaf-2015-0506>.
- Schindler, Kirsten. 2021. «Literarisches Übersetzen – Eine besondere Form des Schreibens». *trans-kom* 14 (1): 43–62.
- Schmälting, Benjamin. 2018. «Dhoch3: Online-Studienmodule für die Deutschlehrausbildung an Hochschulen weltweit: kulturspezifische Modifizierbarkeit und Anwendung im Blended-Learning-Format». *Informationen Deutsch als Fremdsprache* 45 (5): 635–54. <https://doi.org/doi:10.1515/infodaf-2018-0083>.
- Schum, Swenia. 2017. «Barrierefreiheit als Herausforderung in der Fachtextübersetzung». *trans-kom* 10 (3): 349–63.
- Seyferth-Zapf, Maria, e Silke Grafe. 2022. «Gemeinsame Gestaltung medialer Beiträge als Ziel interkultureller Begegnung». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 21 (35): 87–108. <https://doi.org/10.24403/jp.1295945>.
- Stang, Alexandra. 2022. «Theoretische Reflexionen und Konturen einer relational orientierten Wissenskommunikation. Überlegungen für eine künftige Forschungsagenda». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 21 (35): 47–64. <https://doi.org/10.24403/jp.1295924>.

Wichmann, Martin. 2016. «Jugendsprache in DaF-Lehrwerken – oft gefordert, aber selten praktiziert. Eine Bestandsaufnahme». *Informationen Deutsch als Fremdsprache* 43 (6): 667–92. <https://doi.org/doi:10.1515/infodaf-2016-0608>.

Letteratura critica:

Ageeva, Ekaterina, Mikel L. Forcada, Francis M. Tyers, e Juan Antonio Pérez-Ortiz. 2015. «Evaluating machine translation for assimilation via a gap-filling task». In *Proceedings of the 18th Annual Conference of the European Association for Machine Translation*, a cura di İlknur Durgar El-Kahlout, Mehmed Özkan, Felipe Sánchez-Martínez, Gema Ramírez-Sánchez, Fred Hollowood, e Andy Way. <https://aclanthology.org/W15-4918/>.

Allen, Jeffrey H. 2003. «Post-Editing». In *Computers and Translation: A Translator's Guide*, a cura di Harold Somers, vol. 35. John Benjamins Publishing Company. <https://doi.org/10.1075/btl.35.19all>.

Alonso, Elisa, e Elisa Calvo. 2015. «Developing a Blueprint for a Technology-mediated Approach to Translation Studies». *Meta* 60 (1): 135–57. <https://doi.org/10.7202/1032403ar>.

Alvarez-Vidal, Sergi, Antoni Oliver, e Toni Badia. 2020. «Post-editing for Professional Translators: Cheer or Fear?» *Tradumàtica tecnologies de la traducció*, fasc. 18: 49–69. <https://doi.org/10.5565/rev/tradumatica.275>.

Ammon, Ulrich. 1973. *Probleme der Soziolinguistik*. Germanistische Arbeitshefte 15. Max Niemeyer.

Arenas, Ana Guerberof, e Antonio Toral. 2020. «The Impact of Post-editing and Machine Translation on Creativity and Reading Experience». *Translation Spaces* 9 (2): 255–82. <https://doi.org/10.1075/ts.20035.gue>.

Bahdanau, Dzmitry, Kyunghyun Cho, e Yoshua Bengio. 2016. «Neural Machine Translation by Jointly Learning to Align and Translate». arXiv:1409.0473. Preprint, arXiv. <https://doi.org/10.48550/arXiv.1409.0473>.

Banerjee, Satanjeev, e Alon Lavie. 2005. «METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments». In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, a cura di Jade Goldstein, Alon Lavie, Chin-Yew Lin, e Clare Voss. Association for Computational Linguistics. <https://aclanthology.org/W05-0909/>.

Bar-Hillel, Yehoshua. 1960. «The Present Status of Automatic Translation of Languages». In *Advances in Computers*, vol. 1. Elsevier. [https://doi.org/10.1016/S0065-2458\(08\)60607-5](https://doi.org/10.1016/S0065-2458(08)60607-5).

Bentivogli, Luisa, Mauro Cettolo, Marcello Federico, e Christian Federmann. 2018. «Machine Translation Human Evaluation: an investigation of evaluation based on Post-Editing and its relation with Direct Assessment». In *Proceedings of the 15th International Conference on Spoken Language Translation*, a cura di Marco Turchi, Jan Niehues, e Marcello Federico. International Conference on Spoken Language Translation. <https://aclanthology.org/2018.iwslt-1.9/>.

- Bertoldi, Nicola, Davide Caroselli, e Marcello Federico. 2018. «The ModernMT Project». In *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, a cura di Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez, Miquel Esplà-Gomis, et al. <https://aclanthology.org/2018.eamt-main.46/>.
- Bertoldi, Nicola, e Marcello Federico. 2009. «Domain adaptation for statistical machine translation with monolingual resources». *Proceedings of the Fourth Workshop on Statistical Machine Translation (USA)*, StatMT '09, 182–89.
- Bianchi, Marco. 2020. *Galileo in Europa: La scelta del volgare e la traduzione latina del Dialogo sopra i due massimi sistemi*. Filologie medievali e moderne 5. Fondazione Università Ca' Foscari.
- Birch, Alexandra, Omri Abend, Ondřej Bojar, e Barry Haddow. 2016. «HUME: Human UCCA-Based Evaluation of Machine Translation». In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, a cura di Jian Su, Kevin Duh, e Xavier Carreras. Association for Computational Linguistics. <https://doi.org/10.18653/v1/D16-1134>.
- Blatz, John, Erin Fitzgerald, George Foster, et al. 2004. «Confidence Estimation for Machine Translation». *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, 315–21. <https://aclanthology.org/C04-1046/>.
- Bojar, Ondřej, Rajen Chatterjee, Christian Federmann, et al. 2017. «Findings of the 2017 Conference on Machine Translation (WMT17)». In *Proceedings of the Second Conference on Machine Translation*, a cura di Ondřej Bojar, Christian Buck, Rajen Chatterjee, et al. Association for Computational Linguistics. <https://doi.org/10.18653/v1/W17-4717>.
- Bojar, Ondřej, Christian Federmann, Barry Haddow, Philipp Koehn, Matt Post, e Lucia Specia. 2016. «Ten Years of WMT Evaluation Campaigns: Lessons Learnt». *Proceedings of the LREC 2016 Workshop “Translation Evaluation – From Fragmented Tools and Data Sets to an Integrated Ecosystem”*, 27–34. <https://www.research.ed.ac.uk/en/publications/ten-years-of-wmt-evaluation-campaigns-lessons-learnt>.
- Bojar, Ondřej, e Aleš Tamchyna. 2011. «Improving Translation Model by Monolingual Data». In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, a cura di Chris Callison-Burch, Philipp Koehn, Christof Monz, e Omar F. Zaidan. Association for Computational Linguistics. <https://aclanthology.org/W11-2138/>.
- Bowker, Lynne. 2002. *Computer-aided translation technology: a practical introduction*. Didactics of translation series. University of Ottawa Press.
- Bowker, Lynne. 2019. «Fit-for-Purpose Translation». In *The Routledge Handbook of Translation and Technology*, a cura di Minako O'Hagan. Routledge Handbooks in Translation and Interpreting Studies. Routledge. <https://doi.org/10.4324/9781315311258>.
- Brown, Peter F., John Cocke, Stephen Della Pietra, et al. 1988. «A Statistical Approach to Language Translation». Atto di convegno presented a COLING 1988. *Coling Budapest 1988 Volume 1: International Conference on Computational Linguistics*. <https://aclanthology.org/C88-1016/>.

- Brown, Tom B., Benjamin Mann, Nick Ryder, et al. 2020. «Language models are few-shot learners». *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Red Hook, NY, USA), NIPS '20.
- Bucchi, Massimiano. 2000. *La scienza in pubblico: percorsi nella comunicazione scientifica*. McGraw-Hill.
- Cadwell, Patrick, Sharon O'Brien, e Carlos S. C. Teixeira. 2018. «Resistance and Accommodation: Factors for the (Non-) Adoption of Machine Translation among Professional Translators». *Perspectives* 26 (3): 301–21. <https://doi.org/10.1080/0907676X.2017.1337210>.
- Calaresu, Emilia Maria. 2006. «L'universalità del linguaggio scientifico tra norma d'uso e sistema linguistico. Monolinguismo e plurilinguismo nella comunicazione scientifica». In *Italienisch und Deutsch als Wissenschaftssprachen. Bestandsaufnahmen, Analysen, Perspektiven // Italiano e tedesco come lingue della comunicazione scientifica. Riconoscizioni, analisi, prospettive*, a cura di Emilia Maria Calaresu, Cristina Guardiano, e Klaus Hölker. Lit Verlag. <https://iris.unimore.it/handle/11380/461173?mode=simple>.
- Callison-Burch, Chris, Cameron Fordyce, Philipp Koehn, Christof Monz, e Josh Schroeder. 2007. «(Meta-) Evaluation of Machine Translation». In *Proceedings of the Second Workshop on Statistical Machine Translation*, a cura di Chris Callison-Burch, Philipp Koehn, Cameron Shaw Fordyce, e Christof Monz. Association for Computational Linguistics. <https://aclanthology.org/W07-0718/>.
- Castilho, Sheila, Stephen Doherty, Federico Gaspari, e Joss Moorkens. 2018. «Approaches to Human and Machine Translation Quality Assessment». In *Translation Quality Assessment: From Principles to Practice*, a cura di Joss Moorkens, Sheila Castilho, Federico Gaspari, e Stephen Doherty. ProQuest Ebook Central, volume 1. Springer.
- Chatterjee, Rajen, Matteo Negri, Marco Turchi, Marcello Federico, Lucia Specia, e Frédéric Blain. 2017. «Guiding Neural Machine Translation Decoding with External Knowledge». In *Proceedings of the Second Conference on Machine Translation*, a cura di Ondřej Bojar, Christian Buck, Rajen Chatterjee, et al. Association for Computational Linguistics. <https://doi.org/10.18653/v1/W17-4716>.
- Chatzikoumi, Eirini. 2020. «How to Evaluate Machine Translation: A Review of Automated and Human Metrics». *Natural Language Engineering* 26 (2): 137–61. <https://doi.org/10.1017/S1351324919000469>.
- Chen, Mark, Jerry Tworek, Heewoo Jun, et al. 2021. «Evaluating Large Language Models Trained on Code». arXiv:2107.03374. Preprint, arXiv, luglio 14. <https://doi.org/10.48550/arXiv.2107.03374>.
- Chiocchetti, Elena, e Flavia De Camillis. 2022. «MT@BZ Annotation Guidelines v1.0». *Eurac Research CLARIN Centre*. <http://hdl.handle.net/20.500.12124/62>.
- Cho, Kyunghyun, Bart van Merriënboer, Dzmitry Bahdanau, e Yoshua Bengio. 2014. «On the Properties of Neural Machine Translation: Encoder–Decoder Approaches». In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, a

- cura di Dekai Wu, Marine Carpuat, Xavier Carreras, e Eva Maria Vecchi. Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-4012>.
- Cho, Kyunghyun, Bart van Merriënboer, Caglar Gulcehre, et al. 2014. «Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation». *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724–34. <https://doi.org/10.3115/v1/D14-1179>.
- Cloître, Michel, e Terry Shinn. 1985. «Expository Practice: social, cognitive and epistemological linkages». In *Expository Science: Forms and Functions of Popularisation*, edited by Richard D. Whitley, vol. 9, a cura di Terry Shinn e Richard D. Whitley. Sociology of the Sciences a Yearbook. Springer Netherlands. https://doi.org/10.1007/978-94-009-5239-3_2.
- Cortelazzo, Michele A. 1994. «Testo scientifico e manuali scolastici». In *La rete e i nodi: il testo scientifico nella scuola di base*, 1. ed, a cura di Maria Luisa Zambelli. Didattica viva 220. La Nuova Italia.
- Costa, Ângela, Wang Ling, Tiago Luís, Rui Correia, e Luísa Coheur. 2015. «A Linguistically Motivated Taxonomy for Machine Translation Error Analysis». *Machine Translation* 29 (2): 127–61. <https://doi.org/10.1007/s10590-015-9169-0>.
- Costa, Marcella. 2011. «Interpunktionspraktiken im Vergleich am Beispiel von Punkt und Gedankenstrich». In *Deutsch - Italienisch: Sprachvergleiche, Tedesco-Italiano: confronti linguistici*, a cura di Maria Sandra Coletos Bosco, Marcella Costa, Ludwig M. Eichinger, Maria Sandra Coletos Bosco, e Ludwig M. Eichinger. Germanistische Bibliothek 40. Internationale Tagung Tedesco - Italiano: Confronti Linguistici, Deutsch - Italienisch: Sprachvergleiche, Heidelberg. Winter.
- Crestani, Valentina. 2022. «Il “linguaggio facile tedesco”, il “linguaggio facile italiano”: una prospettiva sulle strategie di mediazione». *ItalianoLinguaDue* 14: 237–59. <https://dx.doi.org/10.54103/2037-3597/19613>.
- Crestani, Valentina. 2024. «“Deutsche Leichte Sprache” e “linguaggio facile italiano”: un confronto introduttivo». In *Il tedesco tra lingua difficile e “lingua facile”: Prospettive sulla Leichte Sprache*, a cura di Claudio Di Meola, Daniela Puato, e Ciro Porcaro. Sapienza Università Editrice. <https://air.unimi.it/handle/2434/1079748>.
- Daems, Joke, Lieve Macken, e Sonia Vandepitte. 2013. *Annotation Guidelines for English-Dutch Translation Quality Assessment, version 1.1*. LT3 Technical Report Series. <https://lt3.ugent.be/publications/annotation-guidelines-for-english-dutch-translation/>.
- Dardano, Maurizio. 1994. «I linguaggi scientifici». In *Storia della lingua italiana*, a cura di Luca Serianni e Pietro Trifone, vol. 2. Scritto e parlato. Einaudi.
- DePalma, Donald A. 2017a. «Augmented Translation Powers up Language Services». *Our Analysts' Insights - CSA Research*, febbraio 15. <https://csa-research.com/Blogs-Events/Blog/Augmented-Translation-Powers-up-Language-Services> (ultima consultazione: 20/12/2025).

- DePalma, Donald A. 2017b. «Augmented Translation Powers Up Language Services». CSA Research. <https://csa-research.com/Blogs-Events/Blog/Augmented-Translation-Powers-up-Language-Services> (ultima consultazione: 20/12/2025).
- Dinu, Georgiana, Prashant Mathur, Marcello Federico, e Yaser Al-Onaizan. 2019. «Training Neural Machine Translation to Apply Terminology Constraints». *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3063–68. <https://doi.org/10.18653/v1/p19-1294>.
- Doddington, George. 2002. «Automatic Evaluation of Machine Translation Quality Using N-Gram Co-Occurrence Statistics». *Proceedings of the Second International Conference on Human Language Technology Research*, 128–32. <https://doi.org/10.3115/1289189.1289273>.
- Dorr, Bonnie, Matthew Snover, e Nitin Madnani. 2011. «Chapter 5.1 Introduction». In *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation*, a cura di Joseph Olive, Caitlin Christianson, e John McCary. Springer New York. <https://doi.org/10.1007/978-1-4419-7713-7>.
- Drugan, Joanna, Joss Moorkens, María Fernández-Parra, Andrew Rothwell, e Frank Austermeuhl. 2023. *Translation Tools and Technologies*. 1^a ed. Routledge. <https://doi.org/10.4324/9781003160793>.
- Duden, Richtiges und gutes Deutsch: Wörterbuch der sprachlichen Zweifelsfälle auf der Grundlage der neuen amtlichen Rechtschreibregeln*. 2001. 5., neu Bearb. Aufl. Der Duden in zwölf Bänden 9. Dudenverl.
- Edmundson, Harold P., e David G. Hayes. 1985. «Research methodology for machine translation». *Machine Translation* 5 (1): 8–15.
- Ehlich, Konrad. 1993. «Deutsch als fremde Wissenschaftssprache.» *Jahrbuch Deutsch als Fremdsprache* 19: 13–42.
- Ehlich, Konrad. 1995. «Die Lehre der deutschen Wissenschaftssprache: sprachliche Strukturen, didaktische Desiderate». In *Linguistik der Wissenschaftssprache*, a cura di Heinz Leonhard Kretzenbacher e Harald Weinrich. Akademie der Wissenschaften zu Berlin. Forschungsberichte 10. Walter de Gruyter.
- Ehlich, Konrad. 1996/2007. «Sprache als System versus Sprache als Handlung.» In *Sprache und sprachliches Handeln*, di Konrad Ehlich. Sprache und sprachliches Handeln, Bd. 1. De Gruyter.
- Ehlich, Konrad. 1999. «Alltägliche Wissenschaftssprache». *Informationen Deutsch als Fremdsprache* 26 (1): 3–24. <https://doi.org/10.1515/infodaf-1999-0102>.
- Ehlich, Konrad. 2000. «Deutsch als Wissenschaftssprache für das 21. Jahrhundert». *German as a Foreign Language (GFL)* 1: 47–63.
- ELIS 2025 - European Language Industry Survey*. 2025. https://elis-survey.org/wp-content/uploads/2025/03/ELIS-2025_Report.pdf (ultima consultazione: 20/12/2025).

- Farrús, Mireia, Marta R. Costa-jussà, José B. Mariño, et al. 2011. «Overcoming Statistical Machine Translation Limitations: Error Analysis and Proposed Solutions for the Catalan–Spanish Language Pair». *Language Resources and Evaluation* 45 (2): 181–208. <https://doi.org/10.1007/s10579-011-9137-0>.
- Federico, Marcello, Nicola Bertoldi, Marco Trombetti, e Alessandro Cattelan. 2014. «MateCat: an open source CAT tool for MT post-editing». Atto di convegno presented a Proceedings of the 11th Conference of the Association for Machine Translation in the Americas: Tutorials. <https://aclanthology.org/2014.amta-tutorials.3/>.
- Federmann, Christian. 2010. «Appraise: An Open-Source Toolkit for Manual Phrase-Based Evaluation of Translations». In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, a cura di Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, et al. European Language Resources Association (ELRA). <https://aclanthology.org/L10-1133/>.
- Federmann, Christian. 2012. «Appraise: an Open-Source Toolkit for Manual Evaluation of MT Output». *The Prague Bulletin of Mathematical Linguistics* 98 (1). <https://doi.org/10.2478/v10108-012-0006-9>.
- Feilke, Helmuth. 1998. «Idiomatische Prägung». In *Zwischen Grammatik und Lexikon*, di Irmhild Barz e Günther öhlschläger. Linguistische Arbeiten 390. M. Niemeyer.
- Feuerriegel, Stefan, Jochen Hartmann, Christian Janiesch, e Patrick Zschech. 2024. «Generative AI». *Business & Information Systems Engineering* 66 (1): 111–26. <https://doi.org/10.1007/s12599-023-00834-7>.
- Finn, Chelsea, Pieter Abbeel, e Sergey Levine. 2017. «Model-agnostic meta-learning for fast adaptation of deep networks». *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (Sydney, NSW, Australia), ICML'17, 1126–35.
- Flanagan, Mary. 1994. «Error Classification for MT Evaluation». *Proceedings of the First Conference of the Association for Machine Translation in the Americas*, ottobre 5, 65–72. <https://aclanthology.org/1994.amta-1.9/>.
- Forcada, Mikel L. 2017. «Making Sense of Neural Machine Translation». *Translation Spaces* 6 (2): 291–309. <https://doi.org/10.1075/ts.6.2.06for>.
- Franklin, Michael, e Stan Zdonik. 1998. «“Data in Your Face”: Push Technology in Perspective». *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*, 516–19. <https://doi.org/10.1145/276304.276360>.
- Galilei, Galileo. 1638. *Discorsi e dimostrazioni intorno a due nuove scienze*. Leiden.
- Gandrabur, Simona, e George Foster. 2003. «Confidence estimation for translation prediction». *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, 95–102. <https://aclanthology.org/W03-0413/>.

- Gaspari, Federico. 2011. «Introduzione alla traduzione automatica». In *Usare la traduzione automatica*, a cura di Gabriele Bersani Berselli e Federico Gaspari. Scrittura traduzione tecnologie 1. CLUEB.
- Gaspari, Federico, Antonio Toral, Sudip Kumar Naskar, Declan Groves, e Andy Way. 2014. «Perception vs. reality: measuring machine translation post-editing productivity». In *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas*, a cura di Sharon O'Brien, Michel Simard, e Lucia Specia. Association for Machine Translation in the Americas. <https://aclanthology.org/2014.amta-wptp.5/>.
- Gene, Viveta. 2021. «The Post-Editing Workflow: Training Challenges for LSPs, Post-Editors and Academia». In *Proceedings of the Translation and Interpreting Technology Online Conference*, a cura di Ruslan Mitkov, Vilemini Sosoni, Julie Christine Giguère, Elena Murgolo, e Elizabeth Deysel. INCOMA Ltd. <https://aclanthology.org/2021.triton-1.22/>.
- Goethe, Johann Wolfgang. 1875. *Goethes Sämmtliche Werke. Vollständige Ausgabe in zehn Bänden*. Vol. 8. Verlag der Cotta'schen Buchhandlung.
- Gonzàlez, Meritxell, e Jesús Giménez. 2014. «Asiya. An open toolkit for automatic machine translation (meta-)evaluation. Technical Manual, version 3.0.» TALP Research Center, LSI Department, Universitat Politècnica de Catalunya.
- Goodfellow, Ian, Yoshua Bengio, e Aaron Courville. 2016. *Deep Learning*. Adaptive Computation and Machine Learning. The MIT press.
- Görög, A. 2014. «Quality evaluation today: the Dynamic Quality Framework». Atto di convegno presented a TC 2014, London, UK. *Proceedings of Translating and the Computer 36*. <https://aclanthology.org/2014.tc-1.21/>.
- Graefen, Gabriele. 1997. *Der wissenschaftliche Artikel: Textart und Textorganisation*. Arbeiten zur Sprachanalyse 27. Peter Lang.
- Graefen, Gabriele. 2001. «Einführung in den Gebrauch der Wissenschaftssprache.» In *Wie schwer ist die deutsche Sprache wirklich?*, 1. Aufl, a cura di Armin Wolff e Elmar Winters-Ohle. Materialien Deutsch als Fremdsprache, H. 58. Jahrestagung Deutsch als Fremdsprache, Regensburg. Fachverband Deutsch als Fremdsprache.
- Graefen, Gabriele. 2004. «Aufbau idiomatischer Kenntnisse in der Wissenschaftssprache». In *Integration durch Sprache: Beiträge der 31. Jahrestagung DaF 2003*, 1. Aufl, a cura di Armin Wolff, Torsten Ostermann, e Christoph Chlosta. Materialien Deutsch als Fremdsprache 73. Jahrestagung Deutsch als Fremdsprache, Regensburg. Fachverband Deutsch als Fremdsprache.
- Graefen, Gabriele, e Melanie Moll. 2011. *Wissenschaftssprache Deutsch: lesen – verstehen – schreiben: Ein Lehr- und Arbeitsbuch*. Peter Lang Verlag. <https://doi.org/10.3726/978-3-653-00982-8>.
- Graham, Yvette, Timothy Baldwin, Alistair Moffat, e Justin Zobel. 2017. «Can Machine Translation Systems Be Evaluated by the Crowd Alone». *Natural Language Engineering* 23 (1): 3–30. <https://doi.org/10.1017/S1351324915000339>.

- Graves, Alex. 2012. *Supervised Sequence Labelling with Recurrent Neural Networks*. Vol. 385. Studies in Computational Intelligence. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-24797-2>.
- Graves, Alex, Greg Wayne, e Ivo Danihelka. 2014. «Neural Turing Machines». arXiv:1410.5401. Preprint, arXiv, dicembre 10. <https://doi.org/10.48550/arXiv.1410.5401>.
- Guerrero Romeo, Lucía. 2018. «From a Discreet Role to a Co-Star: The Post-Editor Profile Becomes Key in the PEMT Workflow for an Optimal Outcome». *Proceedings of the 40th Conference Translating and the Computer*, 89–94.
- Han, Aaron L. F., Derek F. Wong, e Lidia S. Chao. 2012. «LEPOR: A Robust Evaluation Metric for Machine Translation with Augmented Factors». In *Proceedings of COLING 2012: Posters*, a cura di Martin Kay e Christian Boitet. The COLING 2012 Organizing Committee. <https://aclanthology.org/C12-2044/>.
- Han, Chao. 2020. «Translation Quality Assessment: A Critical Methodological Review». *The Translator* 26 (3): 257–73. <https://doi.org/10.1080/13556509.2020.1834751>.
- Haykin, Simon S. 2009. *Neural Networks and Learning Machines*. 3. ed. Prentice-Hall.
- Heiss, Christine, e Marcello Soffritti. 2018. «DeepL Traduttore e didattica della traduzione dall'italiano in tedesco Alcune valutazioni preliminari». in *TRAlinea: Online Translation Journal Special Issue: Translation And Interpreting for Language Learners (TAIL)*: 241–58.
- Hoffmann, Lothar. 1988. *Vom Fachwort zum Fachtext: Beiträge zur angewandten Linguistik*. Forum für Fachsprachen-Forschung, Bd. 5. G. Narr.
- Hokamp, Chris, e Qun Liu. 2017. «Lexically Constrained Decoding for Sequence Generation Using Grid Beam Search». In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, a cura di Regina Barzilay e Min-Yen Kan. Association for Computational Linguistics. <https://doi.org/10.18653/v1/P17-1141>.
- Hu, Ke, e Patrick Cadwell. 2016. «A Comparative Study of Post-editing Guidelines». *Proceedings of the 19th Annual Conference of the European Association for Machine Translation*, 346–53. <https://aclanthology.org/W16-3420/>.
- Hutchins, W. John. 2003a. «ALPAC: The (In)Famous Report». In *Readings in Machine Translation*, a cura di Sergei Nirenburg, Harold L. Somers, e Yorick A. Wilks. The MIT Press. <https://doi.org/10.7551/mitpress/5779.003.0014>.
- Hutchins, W. John. 2003b. «Commercial Systems: The State of the Art». In *Computers and Translation: A Translator's Guide*, a cura di Harold Somers, vol. 35. John Benjamins Publishing Company. <https://doi.org/10.1075/btl.35.13hut>.
- Hutchins, W. John. 2004. «The Georgetown-IBM Experiment Demonstrated in January 1954». In *Machine Translation: From Real Users to Research*, edited by David Hutchison, Takeo Kanade, Josef Kittler, et al., vol. 3265, a cura di Robert E. Frederking e Kathryn B. Taylor. Lecture Notes in Computer Science. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-30194-3_12.

- Hutchins, W. John, e Harold L. Somers. 1992. *An Introduction to Machine Translation*. Academic Press.
- ISO 18587 - Translation services — Post-editing of machine translation output — Requirements. 2017. International Organization for Standardization (ISO).
- IÜD, Universität Heidelberg. 1996. *Korrekturrichtlinien vom 07.02.1996 auf der Grundlage der Fassung der Überarbeitung der Kommission vom 11.11.1996*. Heidelberg, Germany. <https://www.uni-heidelberg.de/imperia/md/content/fakultaeten/neuphil/iask/sued/studium/korrekturrichtlinien.pdf> (ultima consultazione: 20/12/2025).
- Jia, Yanfang, Michael Carl, e Xiangling Wang. 2019. «Post-Editing Neural Machine Translation versus Phrase-Based Machine Translation for English–Chinese». *Machine Translation* 33 (1–2): 9–29. <https://doi.org/10.1007/s10590-019-09229-6>.
- Jiang, Yuchen, Xiang Li, Hao Luo, Shen Yin, e Okyay Kaynak. 2022. «Quo Vadis Artificial Intelligence?» *Discover Artificial Intelligence* 2 (1): 4. <https://doi.org/10.1007/s44163-022-00022-8>.
- Kalverkämper, Hartwig. 1990. «Gemeinsprache und Fachsprachen Plädoyer für eine integrierende Sichtweise». In *Deutsche Gegenwartssprache. Tendenzen und Perspektiven*, a cura di Gerhard Stickel. Jahrbuch 1989 des Instituts für deutsche Sprache. De Gruyter. <https://doi.org/10.1515/9783110622423-006>.
- Kelleher, John D. 2019. *Deep Learning*. MIT Press.
- Kim, HyunJin, Xiaoyuan Yi, JinYeong Bak, et al. 2025. «The Road to Artificial SuperIntelligence: A Comprehensive Survey of Superalignment». *SuperIntelligence - Robotics - Safety & Alignment* 2 (1). <https://doi.org/10.70777/si.v2i1.13963>.
- Koby, Geoffrey S., Paul Fields, Daryl R. Hague, Arle Lommel, e Alan Melby. 2014. «Defining Translation Quality». *Tradumàtica tecnologies de la traducció*, fasc. 12 (dicembre): 413–20. <https://doi.org/10.5565/rev/tradumatica.76>.
- Koehn, Philipp. 2009. *Statistical Machine Translation*. 1^a ed. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815829>.
- Koehn, Philipp, Franz J. Och, e Daniel Marcu. 2003. «Statistical Phrase-Based Translation». *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 127–33. <https://aclanthology.org/N03-1017/>.
- Koponen, Maarit. 2016. «Machine Translation Post-editing and Effort. Empirical Studies on the Post-editing Process». (Helsinki).
- Koponen, Maarit, e Leena Salmi. 2018. «Post-editing quality: Analysing the correctness and necessity of post-editor corrections». *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 16. <https://doi.org/10.52034/lanstts.v16i0.439>.

- Krings, Hans P. 2001. *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. A cura di Geoffrey S. Koby. Translation Studies 5. Kent State University Pr.
- Krüger, Ralph. 2016a. «Contextualising Computer-Assisted Translation Tools and Modelling Their Usability». *trans-kom* 9 (1): 114–48.
- Krüger, Ralph. 2016b. «Situated LSP Translation from a Cognitive Translational Perspective». *Lebende Sprachen* 61 (2). <https://doi.org/10.1515/les-2016-0014>.
- Krüger, Ralph. 2019. «Augmented translation – eine Bestandsaufnahme des rechnergestützten Fachübersetzungsprozesses». *trans-kom* 12 (1): 142–81.
- Lacruz, Isabel, Gregory M. Shreve, e Erik Angelone. 2012. «Average Pause Ratio as an Indicator of Cognitive Effort in Post-Editing: A Case Study». In *Workshop on Post-Editing Technology and Practice*, a cura di Sharon O'Brien, Michel Simard, e Lucia Specia. Association for Machine Translation in the Americas. <https://aclanthology.org/2012.amta-wptp.3/>.
- Läubli, Samuel, Sheila Castilho, Graham Neubig, Rico Sennrich, Qinlan Shen, e Antonio Toral. 2020. «A Set of Recommendations for Assessing Human-Machine Parity in Language Translation». *Journal of Artificial Intelligence Research* 67. <https://doi.org/10.1613/jair.1.11371>.
- Läubli, Samuel, Rico Sennrich, e Martin Volk. 2018. «Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation». In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, a cura di Ellen Riloff, David Chiang, Julia Hockenmaier, e Jun'ichi Tsujii. Association for Computational Linguistics. <https://doi.org/10.18653/v1/D18-1512>.
- Laviosa-Braithwaite, Sara. 1998. «Universals of Translation». In *Routledge Encyclopedia of Translation Studies*, a cura di Mona Baker e Gabriela Saldanha. Routledge.
- Lecci, Claudia, e Elena Di Bello. 2012. *Usare la traduzione assistita*. Scrittura, traduzione, tecnologie. CLUEB.
- Legg, Shane, e Marcus Hutter. 2007. «A Collection of Definitions of Intelligence». *Proceedings of the 2007 Conference on Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms: Proceedings of the AGI Workshop 2006 (NLD)*, 17–24.
- Lepschy, Anna Laura, e Giulio Lepschy. 2008. «Punteggiatura e linguaggio». In *Storia della punteggiatura in Europa*, di Bice Mortara Garavelli. Enciclopedie del sapere. Laterza.
- Leusch, Gregor, Nicola Ueffing, e Hermann Ney. 2003. «A novel string-to-string distance measure with applications to machine translation evaluation». Atto di convegno presented a MTSummit 2003, New Orleans, USA. *Proceedings of Machine Translation Summit IX: Papers*. <https://aclanthology.org/2003.mtsummit-papers.32/>.
- Levenshtein, Vladimir I. 1966. «Binary codes capable of correcting deletions, insertions, and reversals». *Soviet Physics – Doklady* 10 (8): 707–10.

- Lin, Chin-Yew, e Franz Josef Och. 2004. «Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics». *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, 605–12. <https://doi.org/10.3115/1218955.1219032>.
- Lita, Lucian, Monica Rogati, e Alon Lavie. 2005. «BLANC: Learning Evaluation Metrics for MT». In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, a cura di Raymond Mooney, Chris Brew, Lee-Feng Chien, e Katrin Kirchhoff. Association for Computational Linguistics. <https://aclanthology.org/H05-1093/>.
- Little, Roderick JA. 1993. «Statistical Analysis of Masked Data». *Journal of Official Statistics* (Stockholm) 9 (2): 407. Publicly Available Content Database (1266808565).
- Liu, Jiawei, Songrun Xie, Junhao Wang, Yuxiang Wei, Yifeng Ding, e Lingming Zhang. 2024. «Evaluating Language Models for Efficient Code Generation». arXiv:2408.06450. Preprint, arXiv, agosto 12. <https://doi.org/10.48550/arXiv.2408.06450>.
- Loffler-Laurian, Anne-Marie. 1983. «Typologie des discours scientifiques: deux approches». *Études de Linguistique Appliquée* 51: 8–20.
- Loffler-Laurian, Anne-Marie. 1985. «Traduction Automatique et Style». *Babel. Revue Internationale de La Traduction / International Journal of Translation* 31 (2): 70–76. <https://doi.org/10.1075/babel.31.2.03lof>.
- Lommel, Arle. 2017. «How Augmented Translation Affects the Language Services Industry». CSA Research. <https://csa-research.com/Blogs-Events/Blog/How-Augmented-Translation-Affects-the-Language-Services-Industry> (ultima consultazione: 20/12/2025).
- Lommel, Arle, Aljoscha Burchardt, e Vivien Macketanz. 2017. «Will Neural Technology Drive MT into the Mainstream?» MultiLingual. <https://multilingual.com/articles/will-neural-technology-drive-mt-into-the-mainstream/> (ultima consultazione: 20/12/2025).
- Lommel, Arle, Aljoscha Burchardt, e Hans Uszkoreit. 2013. «Multidimensional quality metrics: a flexible system for assessing translation quality». Atto di convegno presented a TC 2013, London, UK. *Proceedings of Translating and the Computer* 35, novembre 28. <https://aclanthology.org/2013.tc-1.6/>.
- Lommel, Arle, Maja Popovic, e Aljoscha Burchardt. 2014. «Assessing inter-annotator agreement for translation error annotation». Atto di convegno presented a Language Resources and Evaluation Conference, Reykjavik, Iceland. *MTE: Workshop on Automatic and Manual Metrics for Operational Translation Evaluation*.
- Lommel, Arle, Hans Uszkoreit, e Aljoscha Burchardt. 2014. «Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics». *Revista tradumàtica: traducció i tecnologies de la informació i la comunicació* 12: 455–63.
- Luccio, Riccardo. 2014. *Il testo scientifico*. Edizioni Università di Trieste. <http://hdl.handle.net/10077/9832>.

- Luong, Minh-Thang, e Christopher Manning. 2015. «Stanford neural machine translation systems for spoken language domains». In *Proceedings of the 12th International Workshop on Spoken Language Translation: Evaluation Campaign*, a cura di Marcello Federico, Sebastian Stüker, e Jan Niehues. Da Nang, Vietnam. <https://aclanthology.org/2015.iwslt-evaluation.11/>.
- Macken, Lieve, Bram Vanroy, Luca Desmet, e Arda Tezcan. 2022. «Literary translation as a three-stage process: machine translation, post-editing and revision». In *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, a cura di Helena Moniz, Lieve Macken, Andrew Rufener, et al. European Association for Machine Translation. <https://aclanthology.org/2022.eamt-1.13/>.
- Martín-Mor, Adrià. 2018. «Impostazione e struttura di un corso intensivo avanzato di tecnologie per la traduzione». *inTRAlinea: Online Translation Journal* 20. https://www.intralinea.org/archive/article/impostazione_e_struttura_di_un_corso_intensivo_avanzato (ultima consultazione: 20/12/2025).
- Massardo, Isabella, Jaap van der Meer, Sharon O'Brien, Fred Hollowood, Nora Aranberri, e Katrin Drescher. 2016. *MT POST-EDITING GUIDELINES*. TAUS Signature Editions.
- McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, e Claude E. Shannon. 2006. «A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955». *AI Magazine* 27 (4): 4. <https://doi.org/10.1609/aimag.v27i4.1904>.
- Melamed, I. Dan, Ryan Green, e Joseph P. Turian. 2003. «Precision and Recall of Machine Translation». *Companion Volume of the Proceedings of HLT-NAACL 2003 - Short Papers*, 61–63. <https://aclanthology.org/N03-2021/>.
- Migliorini, Bruno. 2001. *Storia della lingua italiana*. 12. ed. Tascabili Bompiani. Saggi tascabili 31. Bompiani.
- Mikolov, Tomas, Kai Chen, Greg Corrado, e Jeffrey Dean. 2013. «Efficient Estimation of Word Representations in Vector Space». arXiv:1301.3781. Preprint, arXiv. <https://doi.org/10.48550/arXiv.1301.3781>.
- Mitchell, Linda, Johann Roturier, e Sharon O'Brien. 2013. «Community-based post-editing of machine-translated content: monolingual vs. bilingual». In *Proceedings of the 2nd Workshop on Post-editing Technology and Practice*, a cura di Sharon O'Brien, Michel Simard, e Lucia Specia. <https://aclanthology.org/2013.mtsummit-wptp.5/>.
- Monti, Johanna. 2014. *Le tecnologie per la traduzione nell'era del cloud computing: stato dell'arte e prospettive future*. <https://unora.unior.it/handle/11574/170110>.
- Monti, Johanna. 2019. *Dalla Zairja alla traduzione automatica - Riflessioni sulla traduzione nell'era digitale*. Loffredo.
- Moorkens, Joss, e Sharon O'Brien. 2017. «Assessing User Interface Needs of Post-Editors of Machine Translation». In *Human Issues in Translation Technology*, a cura di Dorothy Kenny. The IATIS Yearbook 22. Routledge. <https://doi.org/10.4324/9781315648934>.

- Mortara Garavelli, Bice. 2003. *Prontuario di punteggiatura*. Universale Laterza 831. Laterza.
- Moslem, Yasmin, Rejwanul Haque, John D. Kelleher, e Andy Way. 2023. «Adaptive Machine Translation with Large Language Models». In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, a cura di Mary Nurminen, Judith Brenner, Maarit Koponen, et al. European Association for Machine Translation. <https://aclanthology.org/2023.eamt-1.22/>.
- Murphy, Kevin P. 2014. *Machine Learning - A Probabilistic Perspective*. Adaptive Computation and Machine Learning. MIT Press.
- Nagao, Makoto. 1984. «A framework of a mechanical translation between Japanese and English by analogy principle». *Proc. of the International NATO Symposium on Artificial and Human Intelligence (USA)*, 173–80.
- Navarro, Federico, Theresa Lillis, Tiane Donahue, et al. 2022. «Rethinking English as a lingua franca in scientific-academic contexts». In *Journal of English for Research Publication Purposes*, vol. 3. fasc. 1. John Benjamins. <https://doi.org/10.1075/jerpp.21012.nav>.
- Navarro, Gonzalo. 2001. «A guided tour to approximate string matching». *ACM Comput. Surv.* 33 (1): 31–88. <https://doi.org/10.1145/375360.375365>.
- Nießen, Sonja, Franz Josef Och, Gregor Leusch, e Hermann Ney. 2000. «An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research». In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)*, a cura di Maria Gavrilidou, George Carayannis, Stella Markantonatou, Stelios Piperidis, e Gregory Stainhauer. European Language Resources Association (ELRA). <https://aclanthology.org/L00-1210/>.
- Nitzke, Jean, e Silvia Hansen-Schirra. 2021. *A Short Guide to Post-Editing*. Translation and Multilingual Natural Language Processing 16. Freie Universität Berlin.
- Nitzke, Jean, Silvia Hansen-Schirra, e Carmen Canfora. 2019. «Risk management and post-editing competence». *The Journal of Specialised Translation*, fasc. 31: 239–59. <https://doi.org/10.26034/cm.jostrans.2019.185>.
- O'Brien, Sharon. 2005. «Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Translatability». *Machine Translation* 19 (1): 37–58. <https://doi.org/10.1007/s10590-005-2467-1>.
- O'Brien, Sharon. 2006. «Pauses as Indicators of Cognitive Effort in Post-editing Machine Translation Output»: *Across Languages and Cultures* 7 (1): 1–21. <https://doi.org/10.1556/Acr.7.2006.1.1>.
- O'Brien, Sharon. 2011. «Towards Predicting Post-Editing Productivity». *Machine Translation* 25 (3): 197–215. <https://doi.org/10.1007/s10590-011-9096-7>.
- O'Brien, Sharon. 2012. «Translation as Human–Computer Interaction». *Translation Spaces* 1 (agosto): 101–22. <https://doi.org/10.1075/ts.1.05obr>.

- O'Brien, Sharon. 2022. «How to Deal with Errors in Machine Translation: Post-Editing». In *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence*, a cura di Dorothy Kenny. Language Science Press. <https://doi.org/10.5281/ZENODO.6759982>.
- O'Brien, Sharon. 2024. «Human-Centered Augmented Translation: Against Antagonistic Dualisms». *Perspectives* 32 (3): 391–406. <https://doi.org/10.1080/0907676x.2023.2247423>.
- Paepcke, Fritz. 1980. «Textverstehen Und Übersetzen». *Babel. Revue Internationale de La Traduction / International Journal of Translation* 26 (4): 199–204. <https://doi.org/10.1075/babel.26.4.07pae>.
- Papineni, Kishore, Salim Roukos, Todd Ward, e Wei-Jing Zhu. 2002. «Bleu: a Method for Automatic Evaluation of Machine Translation». In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, a cura di Pierre Isabelle, Eugene Charniak, e Dekang Lin. Association for Computational Linguistics. <https://doi.org/10.3115/1073083.1073135>.
- Penning, Dieter. 1995. «Landeskunde als Thema des Deutschunterrichts - fächerübergreifend und/oder fachspezifisch?». *Informationen Deutsch als Fremdsprache* 22 (6): 626–40. <https://doi.org/doi:10.1515/infodaf-1995-220603>.
- Perego, Elisa. 2021. «Easy Language in Italy». In *Handbook of Easy Languages in Europe*, di Camilla Lindholm, a cura di Ulla Vanhatalo. Frank & Timme.
- Pérez-Ortiz, Juan Antonio, Mikel L. Forcada, e Felipe Sánchez-Martínez. 2022. «How neural machine translation works». *Machine translation for everyone: Empowering users in the age of artificial intelligence* 18: 141.
- Peris, Álvaro, e Francisco Casacuberta. 2019. «Online Learning for Effort Reduction in Interactive Neural Machine Translation». *Computer Speech & Language* 58 (novembre): 98–126. <https://doi.org/10.1016/j.csl.2019.04.001>.
- Pielmeier, Hélène, e Paul O'Mara. 2020. «The state of the linguist supply chain». CSA Research. <https://cdn2.hubspot.net/hubfs/4041721/Newsletter/The%20State%20of%20the%20Linguist%20Supply%20Chain%202020.pdf> (ultima consultazione: 20/12/2025).
- Plitt, Mirko, e François Masselot. 2010. «A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context». *The Prague Bulletin of Mathematical Linguistics* 93 (1). <https://doi.org/10.2478/v10108-010-0010-x>.
- Popović, Maja. 2015. «chrF: character n-gram F-score for automatic MT evaluation». In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, a cura di Ondřej Bojar, Rajan Chatterjee, Christian Federmann, et al. Association for Computational Linguistics. <https://doi.org/10.18653/v1/W15-3049>.
- Pustejovsky, James, e Amber Stubbs. 2013. *Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications*. 1. ed. O'Reilly.

- Pym, Anthony. 2014. «Translation Skill-Sets in a Machine-Translation Age». *Meta* 58 (3): 487–503. <https://doi.org/10.7202/1025047ar>.
- Quirk, Christopher B. 2004. «Training a Sentence-Level Machine Translation Confidence Measure». In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, a cura di Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa, e Raquel Silva. European Language Resources Association (ELRA). <https://aclanthology.org/L04-1250/>.
- Rapporto ALPAC. 1966. <https://www.mt-archive.net/50/ALPAC-1966.pdf> (ultima consultazione: 20/12/2025).
- Rein, David, Betty Li Hou, Asa Cooper Stickland, et al. 2023. «GPQA: A Graduate-Level Google-Proof Q&A Benchmark». arXiv:2311.12022. Preprint, arXiv. <https://doi.org/10.48550/arXiv.2311.12022>.
- Reinbothe, Roswitha. 1997. «Landeskunde in der Deutschlehrerausbildung». *Informationen Deutsch als Fremdsprache* 24 (4): 499–513. <https://doi.org/doi:10.1515/infodaf-1997-0406>.
- Reinbothe, Roswitha. 2011. «Geschichte des Deutschen als Wissenschaftssprache im 20. Jahrhundert». In *Wissen schaffen, Wissen kommunizieren: Wissenschaftssprachen in Geschichte und Gegenwart*, a cura di Wieland Eins, Helmut Glück, e Sabine Pretschner. Fremdsprachen in Geschichte und Gegenwart, Bd. 8. Harrassowitz.
- Reinke, Uwe. 2004. *Translation memories: Systeme - Konzepte - linguistische Optimierung*. Saarbrücker Beiträge zur Sprach- und Translationswissenschaft 2. Lang.
- Ricken, Ulrich. 1995. «Zum Thema Christian Wolff und die Wissenschaftssprache der deutschen Aufklärung». In *Linguistik der Wissenschaftssprache*, a cura di Heinz Leonhard Kretzenbacher, Harald Weinrich, e Akademie der Wissenschaften zu Berlin (1987). Forschungsbericht 10. Walter de Gruyter.
- Rico Pérez, Celia. 2024. «Re-thinking Machine Translation Post-Editing Guidelines». *The Journal of Specialised Translation*, fasc. 41: 26–47. <https://doi.org/10.26034/cm.jostrans.2024.4696>.
- Rico Pérez, Celia, e Enrique Torrejón. 2012. *Skills and Profile of the New Role of the Translator as MT Post-Editor*. <https://abacus.universidadeuropea.com/handle/11268/657>.
- Roelcke, Thorsten D. 2010. *Fachsprachen*. 3., neu Bearb. Aufl. Grundlagen der Germanistik 37. Schmidt.
- Rothwell, Andrew, Joss Moorkens, María Fernández-Parra, Joanna Drugan, e Frank Austerlühl. 2023. *Translation Tools and Technologies*. Routledge Introductions to Translation and Interpreting. Routledge. <https://doi.org/10.4324/9781003160793>.
- Sakamoto, Akiko. 2019. «Why Do Many Translators Resist Post-Editing? A Sociological Analysis Using Bourdieu's Concepts». *The Journal of Specialised Translation*, fasc. 31: 201–16. <https://doi.org/10.26034/cm.jostrans.2019.183>.
- Sanders, Gregory, Mark Przybicki, Nitin Madnani, e Matthew Snover. 2011. «Chapter 5.1.2 Human Subjective Judgments». In *Handbook of Natural Language Processing and Machine*

Translation: DARPA Global Autonomous Language Exploitation, a cura di Joseph Olive, Caitlin Christianson, e John McCary. Springer New York. <https://doi.org/10.1007/978-1-4419-7713-7>.

Saunders, Danielle. 2021. «Domain Adaptation and Multi-Domain Adaptation for Neural Machine Translation: A Survey». arXiv.Org. <https://arxiv.org/abs/2104.06951v2>.

Saunders, Danielle. 2022. «Domain Adaptation and Multi-Domain Adaptation for Neural Machine Translation: A Survey». *Journal of Artificial Intelligence Research* 75: 351–424. <https://doi.org/10.1613/jair.1.13566>.

Scansani, Randy, Silvia Bernardini, Adriano Ferraresi, e Luisa Bentivogli. 2019. «Do translator trainees trust machine translation? An experiment on post-editing and revision». In *Proceedings of Machine Translation Summit XVII: Translator, Project and User Tracks*, a cura di Mikel Forcada, Andy Way, John Tinsley, Dimitar Shterionov, Celia Rico, e Federico Gaspari. European Association for Machine Translation. <https://aclanthology.org/W19-6711/>.

Scarpa, Federica. 2008. *La traduzione specializzata: un approccio didattico professionale*. Hoepli.

Schilperoord, Joost. 1996. *It's about time: temporal aspects of cognitive processes in text production*. Utrecht studies in language and communication 6. Rodopi.

Schindler, Kirsten. 2021. «Literarisches Übersetzen – Eine besondere Form des Schreibens». *trans-kom* 14 (1): 43–62.

Schütz, Jörg. 1999. «Deploying the SAE J2450 translation quality metric in MT projects». *Proceedings of Machine Translation Summit VII*, 278–84. <https://aclanthology.org/1999.mtsummit-1.41/>.

Schwan, Ben. 2017. «Maschinenintelligenz: Der Besserübersetzer». MIT Technology Review, agosto 31. <https://www.heise.de/hintergrund/Maschinenintelligenz-Der-Besseruebersetzer-3814856.html> (ultima consultazione: 20/12/2025).

Screen, Ben. 2017. «Machine translation and Welsh: Analysing free statistical machine translation for the professional translation of an under-researched language pair». *The Journal of Specialised Translation*, fasc. 28: 317–44. <https://doi.org/10.26034/cm.jostrans.2017.244>.

Sennrich, Rico, Barry Haddow, e Alexandra Birch. 2016. «Improving Neural Machine Translation Models with Monolingual Data». *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 86–96. <https://doi.org/10.18653/v1/P16-1009>.

Servan, Christophe, Josep Crego, e Jean Senellart. 2016. «Domain specialization: a post-training domain adaptation for Neural Machine Translation». arXiv:1612.06141. Preprint, arXiv. <https://doi.org/10.48550/arXiv.1612.06141>.

Shao, Chenze, Fandong Meng, Jiali Zeng, e Jie Zhou. 2024. «Understanding and Addressing the Under-Translation Problem from the Perspective of Decoding Objective». arXiv:2405.18922. Preprint, arXiv. <https://doi.org/10.48550/arXiv.2405.18922>.

- Shi, Freda, Mirac Suzgun, Markus Freitag, et al. 2022. «Language Models are Multilingual Chain-of-Thought Reasoners». arXiv:2210.03057. Preprint, arXiv, ottobre 6. <https://doi.org/10.48550/arXiv.2210.03057>.
- Snover, Matthew, Bonnie Dorr, Rich Schwartz, Linnea Micciulla, e John Makhoul. 2006. «A Study of Translation Edit Rate with Targeted Human Annotation». *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, agosto 8, 223–31. <https://aclanthology.org/2006.amta-papers.25/>.
- Somers, Harold. 1999. «Review Article: Example-Based Machine Translation». *Machine Translation* 14 (2): 113–57. <https://doi.org/10.1023/A:1008109312730>.
- Somers, Harold. 2003. «1. Introduction». In *Benjamins Translation Library*, a cura di Harold Somers, vol. 35. John Benjamins Publishing Company. <https://doi.org/10.1075/btl.35.04som>.
- Specia, Lucia, Dhvaj Raj, e Marco Turchi. 2010. «Machine Translation Evaluation versus Quality Estimation». *Machine Translation* 24 (1): 39–50. <https://doi.org/10.1007/s10590-010-9077-2>.
- Specia, Lucia, Kashif Shah, Jose G. C. de Souza, e Trevor Cohn. 2013. «QuEst - A translation quality estimation framework». In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, a cura di Miriam Butt e Sarmad Hussain. Association for Computational Linguistics. <https://aclanthology.org/P13-4014/>.
- Stahlheber, Eva M. 1992. «Die Fachtextsorte Zeitschriftenartikel im Deutschen und Address/Article im Amerikanischen: Popularisierungsgrad und Diachronie von Funktionen und Strukturen.» In *Kontrastive Fachsprachenforschung*, a cura di Klaus-Dieter Baumann e Hartwig Kalverkämper. Forum für Fachsprachen-Forschung 20. Narr.
- Stanovsky, Gabriel, Noah A. Smith, e Luke Zettlemoyer. 2019. «Evaluating Gender Bias in Machine Translation». In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, a cura di Anna Korhonen, David Traum, e Lluís Màrquez. Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1164>.
- Stolze, Radegundis. 2013. *Fachübersetzen - Ein Lehrbuch für Theorie und Praxis*. 3., Korrigierte Auflage. Forum für Fachsprachen-Forschung, Bd. 89. Frank & Timme.
- Sun, Shuo, Francisco Guzmán, e Lucia Specia. 2020. «Are we Estimating or Guesstimating Translation Quality?» In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, a cura di Dan Jurafsky, Joyce Chai, Natalie Schluter, e Joel Tetreault. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.558>.
- Teixeira, Carlos S.C. 2014. «Perceived vs. measured performance in the post-editing of suggestions from machine translation and translation memories». In *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas*, a cura di Sharon O'Brien,

- Michel Simard, e Lucia Specia. Association for Machine Translation in the Americas. <https://aclanthology.org/2014.amta-wptp.4/>.
- Tezcan, Arda, Joke Daems, e Lieve Macken. 2019. «When a 'sport' is a person and other issues for NMT of novels». In *Proceedings of the Qualities of Literary Machine Translation*, a cura di James Hadley, Maja Popović, Haithem Afli, e Andy Way. European Association for Machine Translation. <https://aclanthology.org/W19-7306/>.
- Tezcan, Arda, Véronique Hoste, e Lieve Macken. 2017. «SCATE taxonomy and corpus of machine translation errors». In *Trends in E-Tools and Resources for Translators and Interpreters*, a cura di Gloria Corpas Pastor e Isabel Durán-Muñoz, vol. 45. Brill.
- Thielmann, Winfried. 2014. «Wissenschaftssprache(n)». *Zielsprache Deutsch* 41 (2): 3–20.
- Thielmann, Winfried. 2017. «Genuin Wissenschaftssprachliche Strukturen». *Informationen Deutsch Als Fremdsprache* 44 (5): 546–69. <https://doi.org/10.1515/infodaf-2017-0091>.
- Thouin, Benoit. 1981. «The Meteo system». In *Translating and the Computer: Practical experience of machine translation*, a cura di Veronica Lawson. Aslib. <https://aclanthology.org/1981.tc-1.4/>.
- Tiedemann, Jörg, Mikko Aulamo, Daria Bakshandaeva, et al. 2024. «Democratizing neural machine translation with OPUS-MT». *Language Resources and Evaluation* 58 (2): 713–55. <https://doi.org/10.1007/s10579-023-09704-w>.
- Tomás, Jesús, Josep Àngel Mas, e Francisco Casacuberta. 2003. «A Quantitative Method for Machine Translation Evaluation». In *Proceedings of the EACL 2003 Workshop on Evaluation Initiatives in Natural Language Processing: are evaluation methods, metrics and resources reusable?*, a cura di Katerina Pastra. Association for Computational Linguistics. <https://aclanthology.org/W03-2804/>.
- Torrejón, Enrique, e Celia Rico. 2012. «Skills and Profile of the New Role of the Translator as MT Post-editor». *Tradumàtica tecnologies de la traducció*, fasc. 10: 10. <https://doi.org/10.5565/rev/tradumatica.18>.
- Ueffing, Nicola, Gholamreza Haffari, e Anoop Sarkar. 2007. «Semi-Supervised Model Adaptation for Statistical Machine Translation». *Machine Translation* 21 (2): 77–94. <https://doi.org/10.1007/s10590-008-9036-3>.
- Ueffing, Nicola, e Hermann Ney. 2005. «Application of word-level confidence measures in interactive statistical machine translation». Atto di convegno presented a EAMT 2005, Budapest, Hungary. *Proceedings of the 10th EAMT Conference: Practical applications of machine translation*. <https://aclanthology.org/2005.eamt-1.35/>.
- Varga, Dániel, Péter Halácsy, András Kornai, Viktor Nagy, László Németh, e Viktor Trón. 2007. «Parallel Corpora for Medium Density Languages». In *Current Issues in Linguistic Theory*, a cura di Nicolas Nicolov, Kalina Bontcheva, Galia Angelova, e Ruslan Mitkov, vol. 292. John Benjamins Publishing Company. <https://doi.org/10.1075/cilt.292.32var>.

- Vaswani, Ashish, Noam Shazeer, Niki Parmar, et al. 2023. «Attention Is All You Need». arXiv:1706.03762. Preprint, arXiv. <https://doi.org/10.48550/arXiv.1706.03762>.
- Veale, Tony, e Andy Way. 1997. «Gaijin: A Bootstrapping, Template-Driven Approach to Example-Based MT». 239–44.
- Vieira, Lucas Nunes. 2020. «Automation Anxiety and Translators». *Translation Studies* 13 (1): 1–21. <https://doi.org/10.1080/14781700.2018.1543613>.
- Vilar, David, Jia Xu, Luis Fernando D'Haro, e Hermann Ney. 2006. «Error Analysis of Statistical Machine Translation Output». In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, a cura di Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, et al. European Language Resources Association (ELRA). <https://aclanthology.org/L06-1244/>.
- Voita, Elena, Pavel Serdyukov, Rico Sennrich, e Ivan Titov. 2018. «Context-Aware Neural Machine Translation Learns Anaphora Resolution». *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1264–74. <https://doi.org/10.18653/v1/P18-1117>.
- Voita, Elena, David Talbot, Fedor Moiseev, Rico Sennrich, e Ivan Titov. 2019. «Analyzing Multi-Head Self-Attention: Specialized Heads Do the Heavy Lifting, the Rest Can Be Pruned». *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 5797–808. <https://doi.org/10.18653/v1/P19-1580>.
- Wang, Pei. 2019. «On Defining Artificial Intelligence». *Journal of Artificial General Intelligence* 10 (2): 1–37. <https://doi.org/10.2478/jagi-2019-0002>.
- Way, Andy. 2013. «Emerging use-cases for machine translation». Atto di convegno presented a TC 2013, London, UK. *Proceedings of Translating and the Computer 35*, novembre 28. <https://aclanthology.org/2013.tc-1.12/>.
- Way, Andy. 2018. «Quality Expectations of Machine Translation». In *Translation Quality Assessment: From Principles to Practice*, a cura di Joss Moorkens, Sheila Castilho, Federico Gaspari, e Stephen Doherty. ProQuest Ebook Central, volume 1. Springer.
- Weaver, Warren. 1949. «Translation». <https://mt-archive.net/50/Weaver-1949-original.pdf> (ultima consultazione: 20/12/2025).
- Wei, Jason, Xuezhi Wang, Dale Schuurmans, et al. 2023. «Chain-of-Thought Prompting Elicits Reasoning in Large Language Models». arXiv:2201.11903. Preprint, arXiv. <https://doi.org/10.48550/arXiv.2201.11903>.
- Weinrich, Harald. 1990. «Formen der Wissenschaftssprache». In *Wissenschaftssprache und Sprachkultur*. Tutzing Materialien 61.
- Weinrich, Harald. 1995. «Wissenschaftssprache, Sprachkultur und die Einheit der Wissenschaft». In *Linguistik der Wissenschaftssprache*, a cura di Heinz Leonhard Kretzenbacher e Harald Weinrich. Akademie der Wissenschaften zu Berlin. Forschungsberichte 10. Walter de Gruyter.

- White, John S. 1995. «Approaches to black box MT evaluation». Atto di convegno presented a MTSummit 1995, Luxembourg, Luxembourg. *Proceedings of Machine Translation Summit V*, luglio 10. <https://aclanthology.org/1995.mtsummit-1.37/>.
- White, John S. 2003. «How to Evaluate Machine Translation». In *Computers and Translation: A Translator's Guide*, a cura di Harold L. Somers. Benjamins Translation Library 0929-7316, v. 35. J. Benjamins.
- Yamada, Kenji, e Kevin Knight. 2001. «A Syntax-based Statistical Translation Model». *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, luglio, 523–30. <https://doi.org/10.3115/1073012.1073079>.
- Yang, Zonghan, Yong Cheng, Yang Liu, e Maosong Sun. 2019. «Reducing Word Omission Errors in Neural Machine Translation: A Contrastive Learning Approach». In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, a cura di Anna Korhonen, David Traum, e Lluís Màrquez. Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1623>.
- Yenduri, Gokul, Ramalingam Murugan, Govardanan Chemmalar Selvi, et al. 2024. «GPT (Generative Pre-Trained Transformer)— A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions». *IEEE Access* 12: 54608–49. <https://doi.org/10.1109/ACCESS.2024.3389497>.
- Ying, Xue. 2019. «An Overview of Overfitting and its Solutions». *Journal of Physics: Conference Series* 1168 (febbraio): 022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>.
- Zanchetta, Eros. 2011. «Traduzione automatica e traduzione assistita». In *Usare la traduzione automatica*, a cura di Gabriele Bersani Berselli e Federico Gaspari. Scrittura traduzione tecnologie 1. CLUEB.
- Zhang, Ying, Stephan Vogel, e Alex Waibel. 2004. «Interpreting BLEU/NIST Scores: How Much Improvement do We Need to Have a Better System?» In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, a cura di Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa, e Raquel Silva. European Language Resources Association (ELRA). <https://aclanthology.org/L04-1489/>.
- Zhou, Ming, Bo Wang, Shujie Liu, Mu Li, Dongdong Zhang, e Tiejun Zhao. 2008. «Diagnostic Evaluation of Machine Translation Systems Using Automatically Constructed Linguistic Check-Points». In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, a cura di Donia Scott e Hans Uszkoreit. Coling 2008 Organizing Committee. <https://aclanthology.org/C08-1141/>.

Sitografia

ChatGPT: <https://chatgpt.com/> (ultima consultazione: 20/12/2025)

Cos'è una rete neurale? – IBM: <https://www.ibm.com/it-it/think/topics/neural-networks> (ultima consultazione: 20/12/2025)

DALL·E 3: <https://openai.com/index/dall-e-3/> (ultima consultazione: 20/12/2025)

DeepL: <https://www.deepl.com/translator> (ultima consultazione: 20/12/2025)

Dizionario di Tedesco online – il Sansoni Tedesco: https://dizionari.corriere.it/dizionario_tedesco/ (ultima consultazione: 20/12/2025)

Duden Online Wörterbuch: <https://www.duden.de/woerterbuch> (ultima consultazione: 20/12/2025)

Einblick – Duden Online Wörterbuch: <https://www.duden.de/rechtschreibung/Einblick> (ultima consultazione: 20/12/2025)

Feld-, Wald- und Wiesendoktor – Elliptische Kompositareihen: <https://grammis.ids-mannheim.de/fragen/3966> (ultima consultazione: 20/12/2025)

EMT competence framework: https://commission.europa.eu/document/download/b482a2c0-42df-4291-8bf8-923922ddc6e1_en?filename=emt_competence_fw_2022_en.pdf (ultima consultazione: 20/12/2025)

European Association of Machine Translation: <https://eamt.org/> (ultima consultazione: 20/12/2025)

European Master's in Translation - Multilingualism, Translation and Language-Based AI Services: https://translation.ec.europa.eu/get-involved-european-language-activities-and-initiatives/european-masters-translation_en (ultima consultazione: 20/12/2025)

Google Scholar: <https://scholar.google.com/> (ultima consultazione: 20/12/2025)

Helsinki-NLP/opus-mt-de-it – Hugging Face: <https://huggingface.co/Helsinki-NLP/opus-mt-de-it> (ultima consultazione: 20/12/2025)

How do you deploy an NVIDIA DGX SuperPOD with DGX GB200 systems? – Blog DeepL: <https://www.deepl.com/en/blog/tech/how-to-deploy-nvidia-dgx-superpod> (ultima consultazione: 20/12/2025)

Hugging Face: <https://huggingface.co/> (ultima consultazione: 20/12/2025)

Informationen für alle – Europäische Regeln, wie man Informationen leicht lesbar und leicht verständlich macht: https://www.inclusion-europe.eu/wp-content/uploads/2017/06/DE_Information_for_all.pdf (ultima consultazione: 20/12/2025)

Informazioni per tutti – Linee guida europee per rendere l'informazione facile da leggere e da capire per tutti: https://www.inclusion-europe.eu/wp-content/uploads/2017/06/IT_Information_for_all.pdf (ultima consultazione: 20/12/2025)

Informazioni su DeepL Write: <https://www.deepl.com/it/products/write> (ultima consultazione: 22/10/2025)

Informazioni su Trados – RWS: <https://www.trados.com/it/about/> (ultima consultazione: 20/12/2025)

Istruzioni per gli autori – Informationen Deutsch als Fremdsprache (Info DaF): <https://share.google/b4vVHnmAHoNZvNOwK> (ultima consultazione: 20/12/2025)

LF Aligner <https://sourceforge.net/projects/aligner/> (ultima consultazione: 20/12/2025)

Linguee: <https://www.linguee.it/> (ultima consultazione: 20/12/2025)

Master Technologies de la traduction (TeTra): <https://all-metz.univ-lorraine.fr/formation/langues-etrangeres-appliquees/master-technologies-de-la-traduction-tetra> (ultima consultazione: 20/12/2025)

memoQ AGT: <https://www.memoq.com/product/memoq-agt/> (ultima consultazione: 20/12/2025)

Midjourney: <https://www.midjourney.com/home> (ultima consultazione: 20/12/2025)

MSc in Translation Technology: <https://www.dcu.ie/courses/postgraduate/salis/msc-translation-technology> (ultima consultazione: 20/12/2025)

Norme redazionali di “Interculture journal”: <https://journals.uni-jena.de/ij/libraryFiles/downloadPublic/1> (ultima consultazione: 20/12/2025)

NVIDIA DGX GB200: <https://www.nvidia.com/it-it/data-center/dgx-gb200/> (ultima consultazione: 20/12/2025)

OPUS – Corpora: <https://opus.nlpl.eu/> (ultima consultazione: 20/12/2025)

OPUS-CAT MT Engine and CAT plugins: <https://helsinki-nlp.github.io/OPUS-CAT/> (ultima consultazione: 20/12/2025)

Opus MT-train: Models: <https://github.com/Helsinki-NLP/Opus-MT-train/tree/master/models> (ultima consultazione: 20/12/2025)

Piani tariffari di DeepL: <https://www.deepl.com/it/pro> (ultima consultazione: 22/10/2025)

Prima versione di MQM (Multidimensional Quality Metrics): <https://web.archive.org/web/20211216145215/https://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html> (ultima consultazione: 20/12/2025)

Reverso Context: <https://context.reverso.net/traduzione/> (ultima consultazione: 20/12/2025)

Sketch Engine: <https://www.sketchengine.eu/> (ultima consultazione: 20/12/2025)

Stranizzare / stranizzarsi - Consulenza Linguistica - Accademia della Crusca
<https://accademiadellacrusca.it/it/consulenza/stranizzare--stranizzarsi/994> (ultima consultazione: 20/12/2025)

SYSTRAN: <https://www.systransoft.com/systran/> (ultima consultazione: 20/12/2025)

Tatoeba-Challenge: <https://github.com/Helsinki-NLP/Tatoeba-Challenge/tree/master/models> (ultima consultazione: 20/12/2025)

TAUS: <https://www.taus.net/> (ultima consultazione: 20/12/2025)

Transkribus: <https://www.transkribus.org/it> (ultima consultazione: 20/12/2025)

Treccani Vocabolario: <https://www.treccani.it/vocabolario/> (ultima consultazione: 20/12/2025)

Ultima versione di MQM (Multidimensional Quality Metrics): <https://themqm.org/the-mqm-full-typology/> (ultima consultazione: 20/12/2025)

What Is MQM? – MQM (Multidimensional Quality Metrics): <https://themqm.org/> (ultima consultazione: 20/12/2025)

What is Machine Translation? – European Association of Machine Translation: <https://eamt.org/what-is-machine-translation/> (ultima consultazione: 20/12/2025)

Allegati

Allegato 1 – Lista delle categorie di annotazione

Accuracy

Addition	A/Add
Omission	A/Omis
Untranslated	A/Untra
Do-not-translate	A/DNT
Foreign language	A/ForLa
Mistranslation	
Semantically unrelated	A/MisTra/SemUn
Part-of-speech	A/MisTra/POS
Word sense disambiguation	A/MisTra/WSD
Partial	A/MisTra/Part
Gender	A/MisTra/Gend
Wrong prep/conj	A/MisTra/WroPr
Multi-Word-Expression	
Komposita	A/MWE/Komp
Ellipsis	A/MWE/Ell
Terminology	
Semantically unrelated	A/Termi/SemUn
Do-not-translate	A/Termi/DNT
Missed explication	A/MisExpl
Misordering	
Word order	A/MisOrd/WO
Misplaced word	A/MisOrd/MW
Mechanical	
Punctuation	A/Mech/Punct
Capitalization	A/Mech/Capit
Character formatting	A/CharFor
Source Error	A/SouEr
Other Accuracy Error	A/Oth

Fluency

Lexicon

Lexical choice

F/Lex/LexCho

Fügung

F/Lex/Fueg

Scientific linguistic phenomena

Nominalization

F/SLP/Nomin

Impersonal

F/SLP/Imp

Style

Disfluency

F/Sty/Disfl

Register

F/Sty/Reg

Idiomaticity

F/Sty/Idiom

Design

Character formatting

F/De/CharFor

Layout

F/De/Lay

Coherence

Co-reference

F/Coh/CoRef

Inconsistent Punctuation

F/Coh/InPu

Citation

F/Coh/Cit

Grammar

Orthography

Spelling

G/Orth/Spe

Capitalization

G/Orth/Capit

Punctuation

G/Orth/Punct

White space

G/Orth/WhiSpa

Word form

Agreement

G/WF/Agree

Verb form

G/WF/VF

Allegato 2 – Progetto n. 1 *Informationen Deutsch als Fremdsprache*: analisi dei singoli testi

L'analisi comprende una tabella degli errori annotati, accompagnata da un breve commento sulle categorie più significative, le indicazioni temporali delle sessioni di *post-editing* e i risultati del report di Post-Edit Compare presentati mediante una tabella e un breve commento.

Testo n. 1 (6.952 tokens)

Runte, Maren. 2015. «Wie benutzen fortgeschrittene DaF-Lernende Wörterbücher? Eine Eye-Tracking-Studie zur Benutzung von Lernerwörterbüchern und ein Vorschlag zu deren Verbesserung». *Informationen Deutsch als Fremdsprache* 42 (5): 476–98.
<https://doi.org/doi:10.1515/infodaf-2015-0506>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	205
F/Lex/LexCho	160
F/Sty/Disfl	130
A/MisTra/SemUn	79
A/Termi/SemUn	66
A/MisTra/WSD	25
G/Orth/Capit	24
A/Untra	22
A/Omis	18
A/DNT	16
G/Orth/Punct	14
A/MisExpl	14
A/ForLa	13
F/SLP/Nomin	12
F/Sty/Reg	10
G/Orth/Spel	10
A/MisTra/Part	7
G/Orth/WhiSpa	4
A/MisOrd/WO	4
A/Add	4
A/MisTra/WroPr	4
G/WF/Agree	3
A/Mech/Punct	3
A/MisTra/POS	3
A/Termi/DNT	2
F/SLP/ForImp	1

F/Coh/Cit	1
A/Mech/CharFor	1
F/Coh/CoRef	1
F/Sty/Idiom	1
G/WF/VF	1
A/MisOrd/MW	1

Tabella 1 (All. 2): Analisi quantitativa degli errori per il testo n. 1 del progetto n. 1

Nel testo analizzato si riscontrano complessivamente 858 errori, distribuiti tra le tre macrocategorie nel seguente modo: 521 errori di *Fluency*, 282 errori di *Accuracy* e 56 errori di *Grammar*. La predominanza degli errori di *Fluency* è da ricondurre principalmente alle tre categorie più frequenti: F/Coh/InPu con 205 errori, F/Lex/LexCho con 160 e F/Sty/Disfl con 130. La categoria di F/Coh/InPu riguarda l'uso incoerente di segni ortografici quale l'apostrofo e le virgolette. Gli errori di F/Lex/LexCho sono particolarmente rilevanti per la resa del linguaggio scientifico poichè riguardano l'impiego di espressioni adeguate all'ambito specialistico e al genere testuale specifico. Gli errori di F/Sty/Disfl segnalano le anomalie nella scorrevolezza della lingua di arrivo. Altra categoria di errori di rilievo è A/MisTra/SemUn con 79 occorrenze, riguardante gli errori di corrispondenza semantica tra il testo di arrivo e il testo di partenza. Si riscontra anche un elevato numero di occorrenze per la categoria A/Termi/SemUn (66) la quale segnala la traduzione errata di termini. La categoria di A/Termi/DNT, relativa alla traduzione non necessaria di termini, registra solo 2 occorrenze. Per quanto riguarda invece gli errori stilistici legati al linguaggio scientifico, si constata 12 errori di mancato uso delle nominalizzazioni (F/SLP/Nomin), 10 errori di uso del registro inadeguato (F/Sty/Reg) e un solo errore di mancato uso della forma impersonale (F/SLP/ForImp). Inoltre, si riscontra solo un errore di F/Coh/Cit, che riguarda la strutturazione incoerente dei riferimenti bibliografici, elementi essenziali negli articoli scientifici.

Durata delle sessioni di post-editing con Qualitivity

Il *post-editing* è stato articolato in otto sessioni, con durate individuali di 01:25:59, 00:55:00, 05:00:19, 02:12:51, 03:09:06, 02:33:54 e 01:06:57⁸³, per un totale complessivo di 16:24:06⁸⁴.

Report di Post-Edit Compare

Il report di Post-Edit Compare registra 247 segmenti, per un totale di circa 5.605 parole. La maggioranza, 240 segmenti, è stata generata automaticamente dal sistema di TA, mentre solo 7 segmenti risultano immessi come *New*, cioè come traduzione inserita manualmente. Non si rileva

⁸³ Tutte le durate riportate negli allegati sono espresse nel formato ore:minuti:secondi (hh:mm:ss).

⁸⁴ Tutte le sessioni di *post-editing* riportate negli allegati 2, 3 e 4 sono state intervallate ogni 01:30:00/01:50:00 con pause di durata variabile, non conteggiate nel tempo totale di lavoro effettivo.

alcun contributo della memoria di traduzione; pertanto, che l'intero testo di partenza è derivato da DeepL.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	35	719	12,83%
95-99%	29	638	11,38%
85-94%	69	1.835	32,74%
75-84%	40	1.024	18,27%
50-74%	27	685	12,22%
New (0-49%)	47	704	12,56%

Tabella 2 (All.2): Distribuzione delle fasce di PEMp per testo n. 1 del progetto n. 1

Dalla Tabella 17 emerge una distribuzione delle modifiche piuttosto equilibrata tra le diverse fasce di PEMp, con percentuali che si mantengono tutte entro intervalli medio-bassi. Spiccano tuttavia due valori che si discostano dalla linea generale: la fascia 85-94%, che da sola raccoglie il 32,74% delle parole e la fascia 75-84%, con il 18,27%. Questi due gruppi costituiscono quasi la metà del corpus e segnalano interventi di correzioni lievi (85-94%) e medie (75-84%). Le altre categorie presentano percentuali più contenute: la fascia 100%, che indica i casi di totale riuso dell'output della TA, raggiunge il 12,83%; la fascia 95-99% si ferma all'11,38% e corrisponde a correzioni lievi; la fascia 50-74% copre il 12,22% ed è associata a modifiche di entità medio-alta. La categoria *New*, che copre la quota 0-49%, completa l'analisi con il 12,56%, a testimoniare rielaborazioni più complesse fino a totali su una porzione non trascurabile del testo.

Nel complesso, il profilo evidenzia un output che offre una base traducibile, ma raramente sufficiente senza interventi: solo poco più di un decimo del testo è stato mantenuto inalterato, mentre oltre la metà ha richiesto modifiche significative, con una prevalenza di correzioni lievi e medie. La concentrazione nella fascia 85-94% dimostra che, sebbene le proposte della TA si avvicinino spesso a una resa accettabile, il contributo del *post-editing* resta imprescindibile per raggiungere una qualità pubblicabile.

Testo n. 2 (7.064 tokens)

Geist, Barbara. 2018. «Erhebung syntaktischer Fähigkeiten von Kindern mit Deutsch als Zweitsprache: Ein Methodenvergleich.» *Informationen Deutsch als Fremdsprache* 45 (4): 423–43. <https://doi.org/doi:10.1515/infodaf-2018-0068>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	138
A/MisTra/SemUn	109
F/Lex/LexCho	86
A/Termi/SemUn	56
F/Sty/Disfl	55
A/DNT	51
A/SouEr	48
A/Untra	46
A/Omis	19
G/Orth/Capit	16
F/De/CharFor	16
A/MisTra/Part	15
A/MisOrd/WO	15
A/Add	11
A/MisTra/WSD	9
G/Orth/WhiSpa	8
F/Sty/Reg	7
A/MisOrd/MW	7
F/SLP/Nomin	6
G/WF/Agree	6
A/MisExpl	5
F/Lex/Fue	4
A/MisTra/POS	4
A/MisTra/WroPr	4
F/Coh/CoRef	4
G/Orth/Punct	3
A/MWE/Komp	3
A/Mech/Punct	3
A/ForLa	3
G/Orth/Spel	2
F/Coh/Cit	2
A/MisTra/Gend	2
A/Mech/CharFor	1
A/MWE/Ell	1

Tabella 3 (All. 2): Analisi quantitativa degli errori per il testo n. 2 del progetto n. 1

Si identificano complessivamente 765 errori, distribuiti tra le tre macrocategorie considerate: 318 errori di *Fluency*, 412 errori di *Accuracy* e 35 errori di *Grammar*. La macrocategoria *Accuracy* presenta il numero più elevato di occorrenze, risultando superiore alla somma delle altre due macrocategorie. Considerando la distribuzione interna alle categorie, in ordine decrescente di frequenza emergono: F/Coh/InPu con 138 errori, A/MisTra/SemUn con 109 errori e F/Lex/LexCho con 86 errori. La categoria F/Coh/InPu riguarda l'uso incoerente dell'apostrofo e delle virgolette.

La categoria A/MisTra/SemUn segnala gli errori di corrispondenza semantica tra il testo di arrivo e il testo di partenza. Invece, gli errori di F/Lex/LexCho riguardano le espressioni inadeguate all'ambito specialistico e al genere testuale trattati. Per quanto concerne gli errori stilistici legati al linguaggio scientifico, si annotano 7 errori di registro inadeguato (F/Sty/Reg), 6 errori di nominalizzazione (F/SLP/Nomin) e 4 errori di traduzione delle *Fügungen* (F/Lex/Fue). Con riferimento alla terminologia, la categoria A/Termi/SemUn registra 56 casi di traduzione errata di termini. Si riscontrano solo 2 errori di F/Coh/Cit, riferibili alla strutturazione incoerente dei riferimenti bibliografici. Ulteriori categorie particolarmente rilevanti per numero di occorrenze sono le seguenti: F/Sty/Disfl riguardante la scorrevolezza della lingua di arrivo (55 occorrenze); A/DNT per i segmenti da non tradurre (51 occorrenze); A/SouEr per gli errori nel testo di partenza (48 occorrenze); A/Untra per i segmenti non tradotti (46 occorrenze). L'elevata frequenza della categoria A/DNT è attribuibile alla necessità di mantenere in lingua originale determinati enunciati o termini presi in esame nello studio. Gli errori della categoria A/SouEr sono causati in parte dall'errata collocazione delle note a piè di pagina, problematica che si verifica nel file traducibile generato da Trados Studio, e quindi anche nel testo di arrivo, ma non nel file di origine; in altri casi, sono riconducibili a problemi di formattazione generati durante la conversione del testo di partenza in file traducibile.

Durata delle sessioni di post-editing con Qualitivity

Il *post-editing* è stato svolto in cinque sessioni di lavoro, della durata rispettiva di 02:31:12, 03:10:55, 01:40:05, 03:08:30 e 02:25:31, per un totale complessivo di 12:56:13.

Report di Post-Edit Compare

Il testo analizzato comprende complessivamente 359 segmenti, per un totale di circa 5.688 parole. Di questi, 351 segmenti vengono tradotti mediante il sistema di TA DeepL, mentre non si registrano contributi significativi provenienti dalla memoria di traduzione. Ciò evidenzia che la traduzione si è basata esclusivamente sull'output elaborato dal sistema di TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	70	589	10,36%
95-99%	41	972	17,09 %
85-94%	65	1443	25,37%
75-84%	40	793	13,94%
50-74%	44	679	11,94%
New (0-49%)	99	1212	21,31%

Tabella 4 (All. 2): Distribuzione delle fasce di PEMp per testo n. 2 del progetto n. 1

Dalla tabella 19 si possono estrapolare le seguenti percentuali di distribuzione degli interventi. I dati più rilevanti riguardano due fasce in particolare: la fascia 85-94% registra il 25,37% indicando che circa un quarto del testo ha richiesto delle modifiche lievi. Invece, la categoria *New* (0-49%) presenta il 21,31%, dato che evidenzia come il *post-editing* non si sia limitato a correzioni lievi, ma abbia spesso richiesto interventi sostanziali di riscrittura. Il 17,09% ricade nella fascia 95-99%, segnalando che una parte non trascurabile del testo ha richiesto dei micro-interventi estremamente mirati. Le fasce 75-84% e 50-74%, riguardanti gli interventi di entità media, presentano rispettivamente il 13,95% e il 11,94%. Infine, la fascia 100%, che indica i segmenti che non hanno subito alcuna modifica, in questo caso corrisponde al 10,36% delle parole totali (70 segmenti, 589 parole). Si tratta di un risultato particolarmente significativo, poiché dimostra che solo un decimo del testo prodotto dalla TA è stato riutilizzato integralmente, senza alcuna correzione.

In sintesi, il report documenta un'elevata incidenza di modifiche, che conferma come il sistema di TA riesca a produrre un output che però non è immediatamente pubblicabile. Pertanto, l'intervento umano risulta essenziale per assicurare qualità e correttezza, con un *effort* che richiede prevalentemente correzioni di lieve entità, riscritture totali o pressoché totali e micro-correzioni; in misura più limitata richiede correzioni di media e medio-alta entità.

Testo n. 3 (5.652 tokens)

Pust, Daniel. 2019. «Integrationsmechanismen des Blended Learning im Fremdsprachenunterricht. Eine explorative Fallstudie». *Informationen Deutsch als Fremdsprache* 46 (5): 586–601. <https://doi.org/doi:10.1515/infodaf-2018-0077>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	161
F/Sty/Disfl	81
A/MisTra/SemUn	62
F/Lex/LexCho	57
A/DNT	14
A/Untra	13
A/Omis	12
A/MisTra/Part	11
A/Termi/SemUn	10
A/MisTra/WSD	8
G/Orth/Punct	8
G/WF/Agree	8
A/Add	7

F/SLP/Nomin	5
F/Coh/Cit	5
G/Orth/Capit	5
F/Sty/Reg	4
A/MisTra/WroPr	4
A/Termi/DNT	4
A/MisOrd/WO	3
A/Mech/Punct	3
G/WF/VF	3
G/Orth/WhiSpa	2
F/Lex/Fue	2
G/Orth/Spel	2
F/Coh/CoRef	2
A/ForLa	2
A/SouEr	1
F/SLP/ForImp	1
A/Mech/CharFor	1
F/Sty/Idiom	1

Tabella 5 (All. 2): Analisi quantitativa degli errori per il testo n. 3 del progetto n. 1

Nel presente testo si riscontrano 319 errori di *Fluency*, 155 errori di *Accuracy* e 28 errori di *Grammar*. Il totale degli errori annotati ammonta a 502 errori. La macrocategoria *Fluency* risulta quella più rappresentata, a causa dell'elevata frequenza degli errori di F/Coh/InPu (161 occorrenze) riguardanti l'uso incoerente di virgolette e apostrofi. Tra gli errori più frequenti si individuano quelli di scorrevolezza della lingua di arrivo (F/Sty/Disfl, 81 occorrenze) e di mancata corrispondenza semantica tra il testo di partenza e quello di arrivo (A/MisTra/SemUn, 62 occorrenze). In merito agli errori stilistici del linguaggio scientifico, si rilevano 62 errori di scelta lessicale inadeguata al genere testuale e all'ambito trattati (F/Lex/LexCho), 5 errori di mancata nominalizzazione (F/SLP/Nomin), 4 errori di registro inadeguato (F/Sty/Reg), 2 errori di traduzione delle *Fügungen* (F/Lex/Fue) e un errore di mancata forma impersonale. In riferimento alla componente terminologica, si segnalano 10 errori per la categoria A/Termi/SemUn (termini tradotti in modo errato) e 4 errori di A/Termi/DNT (termini da non tradurre). Si registrano inoltre 5 errori per la categoria F/Coh/Cit, riconducibili alla strutturazione incoerente dei riferimenti bibliografici delle citazioni.

Durata delle sessioni di post-editing con Qualityity

Il *post-editing* è stato effettuato in sei sessioni di lavoro, della durata rispettiva di 01:00:35, 02:17:32, 00:27:14, 02:10:06, 00:25:51 e 03:20:59, per un totale complessivo di 09:42:17.

Report di Post-Edit Compare

Il *post-editing* ha riguardato un totale di 225 segmenti (pari a circa 4.820 parole). Secondo il presente report, la quasi totalità del testo è stata prodotta dal sistema di TA, con 221 segmenti generati automaticamente e solo una minima parte introdotta come *nuova* traduzione (4 segmenti). Non si riscontrano proposte provenienti dalla memoria di traduzione. Pertanto, questo aspetto conferma ancora una volta che la base di partenza è interamente costituita dall'output della TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	52	716	14,85%
95-99%	36	911	18,90%
85-94%	59	1.489	30,89%
75-84%	22	560	11,62%
50-74%	11	214	4,44%
New (0-49%)	45	930	19,29%

Tabella 6 (All. 2): Distribuzione delle fasce di PEMp per testo n. 3 del progetto n. 1

In termini di intervento umano, si individuano le seguenti fasce di percentuali di PEMp nella tabella 21. La quota più consistente di parole si concentra nella fascia 85-94%, il 30,89%, equivalente a quasi un terzo del testo. La fascia 95-99%, invece, presenta il 18,90%. Queste due fasce coprono insieme quasi il 50% della totalità del testo, suggerendo che la metà del testo è interessata da correzioni lievi (fascia 85-94%) o minime (fascia 95-99%) ma necessarie. Rilevante è anche il dato attinente alla categoria *New* (0-49%) che copre il 19,29%, quasi un quinto dell'intero testo. Per le fasce 75-84% e 50-74% le quote sono rispettivamente di 11,62% (560 parole, 22 segmenti) e 4,44%, segnalando valori bassi per gli interventi di media e medio-alta entità. Il 14,85% delle parole è rimasto invariato ed è associato alla fascia 100%. Questa percentuale suggerisce che una porzione del testo, non trascurabile, risultava avere già una qualità pubblicabile.

Complessivamente, i risultati evidenziano che il sistema di TA fornisce un output utilizzabile come base, ma che richiede un volume di correzioni non trascurabile. Infatti, la quota di segmenti accettati senza modifiche suggerisce un'affidabilità del sistema di TA limitata a parti ridotte del testo. L'intervento del *post-editor* si traduce spesso in interventi di minima e lieve entità, ma anche in correzioni estese o riscritture integrali. Minore rilevanza assumono le modifiche di entità media e medio-alta.

Testo n. 4 (6.360 tokens)

Burger, Günter. 2019. «Was bewirkt der Einsatz von Spielfilmen im Fremdsprachenunterricht? Die empirische Forschung im Überblick». *Informationen Deutsch als Fremdsprache* 46 (6): 687–713. <https://doi.org/doi:10.1515/infodaf-2018-0078>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	114
A/MisTra/SemUn	112
F/Lex/LexCho	64
F/Sty/Disfl	61
G/Orth/Punct	38
A/Untra	31
A/Termi/SemUn	27
A/Omis	19
A/SouEr	17
G/Orth/Capit	17
F/SLP/Nomin	15
A/Add	14
A/ForLa	13
A/MisTra/Part	12
G/Orth/WhiSpa	10
A/MisTra/WSD	9
A/MisTra/WroPr	8
G/Orth/Spel	6
A/MisTra/POS	6
F/Sty/Reg	5
F/Lex/Fue	4
A/Mech/Punct	4
G/WF/Agree	3
A/Mech/CharFor	2
A/DNT	1
F/SLP/ForImp	1
A/MisOrd/WO	1
F/Coh/CoRef	1
G/WF/VF	1
A/Termi/DNT	1
A/MisOrd/MW	1

Tabella 7 (All. 2): Analisi quantitativa degli errori per il testo n. 4 del progetto n. 1

Nel complesso si annotano 618 errori, suddivisi nelle tre macrocategorie previste: 278 casi di *Accuracy*, 265 di *Fluency* e 75 di *Grammar*. La differenza tra *Accuracy* e *Fluency* risulta contenuta, con una leggera prevalenza della prima. Le tipologie di errore più ricorrenti sono le seguenti: F/Coh/InPu per l'uso incoerente delle virgolette e dell'apostrofo (114 occorrenze), A/MisTra/SemUn per gli errori di mancata corrispondenza semantica (112 occorrenze), F/Lex/LexCho per gli errori di scelta lessicale (64 occorrenze) e F/Sty/Disfl per gli errori di

scorrevolezza della lingua di arrivo (61 occorrenze). Rilevanti sono anche le occorrenze per la categoria grammaticale riguardante l'uso della punteggiatura (G/Orth/Punct, 38 occorrenze) e la categoria di *Accuracy* che segnala i segmenti non tradotti (A/Untra, 31 occorrenze). Gli errori di natura stilistica, che interessano direttamente il linguaggio scientifico, includono 15 casi di mancato uso della nominalizzazione (F/SLP/Nomin), 5 casi di registro inadeguato (F/Sty/Reg), 4 casi di traduzione errata delle *Fügungen* (F/Lex/Fue) e uno di mancato uso della forma impersonale (F/SLP/ForImp). Per quanto riguarda invece la terminologia, si riscontrano 27 errori di traduzione di termini (A/Termi/SemUn) e un solo errore di traduzione non necessaria (A/Termi/DNT).

Durata delle sessioni di post-editing con Quality

Sono state eseguite quattro sessioni di *post-editing* della durata rispettiva di 02:05:56, 04:44:40, 02:17:38 e 02:54:53, per un totale complessivo di 12:03:07.

Report di Post-Edit Compare

Il report qui analizzato riguarda un testo di 209 segmenti, corrispondenti a circa 5157 parole. La grande maggioranza del materiale di partenza proviene dal sistema di TA: 191 segmenti, pari a 4.599 parole, vengono generate automaticamente, mentre solo 18 segmenti (558 parole) si presentano immessi manualmente come *New*. Non sono presenti *matches* derivanti dalla memoria di traduzione.

Fasce di PEMP	Segmenti	Parole	Percentuali
100%	34	481	9,33%
95-99%	11	309	5,99%
85-94%	57	1.686	32,69%
75-84%	41	1.177	22,82%
50-74%	17	360	6,98 %
New (0-49%)	49	1.144	22,18 %

Tabella 8 (All. 2): Distribuzione delle fasce di PEMP per testo n. 4 del progetto n. 1

Gli interventi di *post-editing* sono caratterizzati da una distribuzione eterogenea delle modifiche: essa evidenzia che la porzione più ampia di interventi si concentra nella fascia 85-94% con il 32,69% e in quella 75-84% con il 22,82%, segnalando circa un terzo e oltre un quinto del testo sono stati interessati da correzioni di lieve entità e di media entità. La fascia 95-99% riguarda il 5,99% e si tratta correzioni di entità minima ma in ogni caso inevitabili; mentre la fascia 50-74% copre il 6,98 %, corrispondente a modifiche medio-alte su porzioni significative di testo. I dati di queste due fasce evidenziano una diffusione limitata di questo tipo di interventi. Rilevante è anche la quota di parole classificate come *New* (pari al 22,18 %), che riflette interventi di riscrittura

integrali o semi integrali e che riguarda oltre un quinto del testo. La fascia 100% indica invece i segmenti non modificati: in questo report ammontano al 9,33 del totale. Questo dato suggerisce che solo una parte del testo generato automaticamente, meno di un decimo, è stata riutilizzata integralmente senza alcuna revisione.

Alla luce dei dati illustrati, si può affermare che il sistema di TA è stato in grado di fornire una traduzione qualitativamente adeguata per una porzione molto limitata di testo. La restante parte è stata interessata da interventi di varia intensità, con una prevalenza di correzioni di lieve e media intensità e di rielaborazioni totali o pressoché integrali. Pertanto, il testo generato automaticamente può essere una base di partenza che però necessita in modo imprescindibile del lavoro di *post-editing*.

Testo n. 5 (5.875 tokens)

Schmälting, Benjamin. 2018. «Dhoch3: Online-Studienmodule für die Deutschlehrerausbildung an Hochschulen weltweit: kulturspezifische Modifizierbarkeit und Anwendung im Blended-Learning-Format». *Informationen Deutsch als Fremdsprache* 45 (5): 635–54. <https://doi.org/doi:10.1515/infodaf-2018-0083>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	135
A/MisTra/SemUn	87
F/Sty/Disfl	64
F/Lex/LexCho	52
A/Untra	35
A/Omis	20
A/MisTra/Part	15
F/Coh/Cit	15
A/Add	13
G/Orth/Capit	12
A/SouEr	10
G/Orth/Punct	9
A/MisTra/WSD	8
A/Termi/SemUn	8
F/Lex/Fue	8
A/MisTra/POS	8
A/Mech/Punct	7
A/Termi/DNT	7
G/Orth/WhiSpa	6
F/SLP/Nomin	6
A/MisTra/WroPr	6

G/Orth/Spel	5
F/Coh/CoRef	5
A/DNT	4
A/ForLa	4
F/Sty/Reg	3
A/MWE/Komp	3
A/MisOrd/WO	3
G/WF/Agree	3
A/Mech/CharFor	3
A/MWE/Ell	3
A/MisOrd/MW	3
A/MisTra/Gend	2
F/De/Layo	2
A/MisExpl	1

Tabella 9 (All. 2): Analisi quantitativa degli errori per il testo n. 5 del progetto n. 1

Nella fase di analisi del testo si riscontrano 290 errori di *Fluency*, 250 errori di *Accuracy* e 35 errori di *Grammar*. Il totale degli errori annotati ammonta a 575. La macrocategoria *Fluency* si conferma la più rappresentata, a causa dell'elevato numero di occorrenze (135) per la categoria F/Coh/InPu che segnala l'uso incoerente delle virgolette e dell'apostrofo. Seguono, in ordine decrescente, le categorie afferenti agli errori di mancata corrispondenza semantica (A/MisTra/SemUn, 87 errori), di scorrevolezza della lingua di arrivo (F/Sty/Disfl, 64 errori) e di scelta lessicale inadeguata all'ambito e al genere testuale trattato (F/Lex/LexCho, 52 errori). Rilevante è anche il numero di annotazioni per la categoria A/Untra, riguardante i segmenti per i quali il sistema di TA non ha proposto una traduzione. In riferimento alle categorie rilevanti per il linguaggio scientifico, si registrano 15 errori di F/Coh/Cit derivanti da una strutturazione incoerente dei riferimenti bibliografici delle citazioni. Invece, per gli aspetti stilistici del linguaggio scientifico, si annotano 8 errori di traduzione delle *Fügungen* (F/Lex/Fue), 6 errori di nominalizzazione (F/SLP/Nomin), 3 errori di registro (F/Sty/Reg). Per la componente terminologica si segnalano 8 errori per la categoria A/Termi/SemUn, riguardante la traduzione errata, e 7 errori per A/Termi/DNT, relativa alla traduzione non necessaria.

Durata delle sessioni di post-editing con Qualitivity

L'attività di *post-editing* è stata articolata in otto sessioni, della durata rispettiva di 02:35:45, 00:58:51, 02:20:22, 01:05:46, 03:13:08 e 02:43:53, per un totale complessivo di 12:57:45.

Report di Post-Edit Compare

Il report di Post-Edit Compare relativo a questo articolo scientifico offre un quadro chiaro della portata del *post-editing* necessario. Il testo analizzato comprende complessivamente 254 segmenti, per un totale di circa 4.975 parole. Di questi segmenti, 239 vengono tradotti mediante il sistema di TA DeepL. L'assenza di *matches*, derivanti dall'impiego della memoria di traduzione, conferma che il testo di partenza è stato interamente prodotto automaticamente, senza alcun contributo da risorse pregresse.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	41	562	11,30%
95-99%	28	515	10,35%
85-94%	60	1.591	31,98%
75-84%	27	582	11,70%
50-74%	16	314	6,31%
New (0-49%)	82	1.411	28,36%

Tabella 10 (All. 2): Distribuzione delle fasce di PEMp per testo n. 5 del progetto n. 1

L'analisi della tabella rivela una distribuzione interessante degli interventi di *post-editing*. Solo l'11,30% delle parole appartiene alla fascia 100%, ovvero porzioni dell'output dell'TA riutilizzate integralmente senza correzioni. Questo dato, pur non trascurabile, evidenzia che la piena affidabilità dell'output automatico resta limitata. La parte più consistente delle modifiche si concentra nella fascia 85-94% dove si registra il 31,98%, segnalando che quasi un terzo del testo è stato interessato da modifiche di lieve entità. Per le fasce 95-99%, 75-84% e 50-74% si osservano dei valori ampiamente contenuti, rispettivamente il 10,35%, l'11,70% e il 6,31%. Tali dati evidenziano una diffusione limitata di interventi di minima, media e medio-alta entità. Invece, un considerevole 28,36% appartiene alla categoria *New* (0-49%), questo significa che poco più di un quarto del testo è stato sottoposto a riscritture complete o quasi, a riprova dell'impatto significativo del *post-editing*.

La distribuzione complessiva suggerisce che il sistema di TA ha formulato delle proposte di traduzione adeguate alla finalità della pubblicazione per una porzione di testo limitata. È stato necessario intervenire sulle altre proposte con correzioni di varia entità, prediligendo in particolare interventi di lieve entità o rielaborazioni integrali o semi-integrali. Sono state apportate anche modifiche di entità minima, media e medio-alta, ma in modo meno diffuso. In sintesi, l'output della TA costituisce un punto di partenza utile ma non è pubblicabile senza l'intervento umano.

Testo n. 6 (8.454 tokens)

Brunsing, Theresa. 2016. «Landeskundliche Abbildungen in Lehrbüchern für Deutsch als Fremdsprache. Die Entwicklung des Bildeinsatzes in den Jahren 2000 bis 2010». *Informationen Deutsch als Fremdsprache* 43 (5): 494–515. <https://doi.org/doi:10.1515/infodaf-2016-0503>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	201
F/Lex/LexCho	97
F/Sty/Disfl	87
A/MisTra/SemUn	74
A/Termi/SemUn	60
A/MisTra/WSD	23
F/SLP/Nomin	21
F/Sty/Reg	19
G/Orth/Capit	19
A/Omis	18
A/DNT	17
A/ForLa	17
A/Untra	14
G/Orth/WhiSpa	13
A/MisTra/Part	13
A/MisOrd/WO	13
A/Termi/DNT	12
G/Orth/Punct	11
G/WF/Agree	11
F/Lex/Fue	10
A/Add	8
F/Coh/Cit	8
G/Orth/Spel	6
A/MisTra/POS	5
A/MisTra/WroPr	5
F/SLP/ForImp	3
A/Mech/Punct	3
F/De/Layo	3
A/MWE/Ell	2
A/MisExpl	2
F/Coh/CoRef	2
A/Mech/CharFor	1
A/SouEr	1
A/MisTra/Gend	1

Tabella 11 (All. 2): Analisi quantitativa degli errori per il testo n. 6 del progetto n. 1

Nel testo analizzato si annotano complessivamente 800 errori, distribuiti tra le tre macrocategorie: 451 di *Fluency*, 289 di *Accuracy* e 60 di *Grammar*. La predominanza degli errori di *Fluency*, che costituiscono più della metà del totale, è principalmente dovuta all'elevato numero di casi di uso incoerente delle virgolette e dell'apostrofo (categoria F/Coh/InPu, 201 errori). Seguono, in ordine decrescente, gli errori di scelta lessicale (F/Lex/LexCho, 97 errori), di scorrevolezza della lingua di arrivo (F/Sty/Disfl, 87 errori) e di mancata corrispondenza semantica (A/MisTra/SemUn, 74 errori). Per gli errori stilistici riguardanti il linguaggio scientifico, si constatano 21 errori di mancata nominalizzazione (F/SLP/Nomin), 19 errori di registro inadeguato (F/Sty/Reg), 10 errori di traduzione delle *Fügungen* (F/Lex/Fue) e 3 errori di mancato uso della forma impersonale (F/SLP/ForImp). La componente terminologia registra 60 errori di traduzione errata (A/Termi/SemUn) e 12 errori di traduzione non necessaria (A/Termi/DNT). Infine, si registrano 8 errori di F/Coh/Cit che segnalano una strutturazione incoerente dei riferimenti bibliografici.

Durata delle sessioni di post-editing con Qualityity

L'attività di *post-editing* è stata distribuita su nove sessioni, con durate rispettive di 02:04:27, 01:06:31, 00:59:33, 01:44:59, 00:56:46, 03:07:16, 00:47:18, 01:11:11 e 01:36:08, per un totale complessivo di 13:34:09.

Report di Post-Edit Compare

Il *post-editing* ha riguardato un totale di 375 segmenti (pari a 6970 parole). Dal report di Post-Edit Compare emerge che 364 segmenti vengono generati automaticamente e solo una minima parte introdotta come 'nuova' traduzione (11 segmenti). Non si riscontrano proposte provenienti dalla memoria di traduzione, in quanto non si registrano *matches*: l'intero testo di partenza deriva dunque dall'output del sistema di TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	75	1.164	16,70%
95-99%	47	933	13,39 %
85-94%	61	1.346	19,31%
75-84%	51	1.158	16,61%
50-74%	45	813	11,94%
New (0-49%)	96	1.556	22,32%

Tabella 12 (All. 2): Distribuzione delle fasce di PEMp per testo n. 6 del progetto n. 1

L'analisi delle distribuzioni delle fasce di PEMp, presentate nella Tabella 27, mette in luce un impegno di correzione distribuito in modo piuttosto omogeneo, ma con aree che spiccano per rilevanza. Il 16,70% delle ricade nella fascia 100%, ovvero dei segmenti mantenuti integralmente

senza alcuna correzione. Questo valore, superiore a quello registrato nei testi finora analizzati, indica che una parte significativa dell'output era già pienamente accettabile. Tuttavia, la restante quota richiede delle modifiche. L'analisi delle restanti fasce di PEMp presenta i seguenti dati: il 13,39 % ricade nella fascia 95-99%; il 19,31% nella fascia 85-94%; il 16,61% nella fascia 75-84%; l'11,94% nella fascia 50-74%. Questi dati evidenziano come il *post-editing* non si sia limitato a correzioni minime o lievi (fasce 95-99% e 85-94%), ma abbia spesso richiesto interventi sostanziali sul testo (fasce 75-84% e 50-74%). Il dato più significativo si registra con la categoria *New* (0-49%) che presenta un consistente 22,32%, a indicare che oltre un quinto del testo ha richiesto riscritture complete o quasi complete.

Complessivamente, il profilo che emerge è quello di un output che fornisce una base disomogenea: alcuni segmenti risultano immediatamente utilizzabili, mentre altri richiedono interventi radicali. Il valore relativamente alto della fascia 100% mostra che la macchina è in grado di produrre, almeno in parte, traduzioni di buona qualità. Tuttavia, la presenza del 22% di segmenti riscritti integralmente, o quasi, riequilibra il giudizio, mettendo in evidenza che la piena affidabilità non può essere data per scontata. In termini di *effort*, il lavoro del *post-editor* spazia quindi da semplici ritocchi fino a rielaborazioni più complesse ed estese.

Testo n. 7 (4.936 tokens)

Funk, Hermann, Manja Gerlach, e Dorothea Spaniel. 2016. «Materialeinsatz in tandembasierten Fremdsprachenlernszenarien. Bericht aus einem laufenden Forschungsprojekt». *Informationen Deutsch als Fremdsprache* 43 (6): 693–709. <https://doi.org/doi:10.1515/infodaf-2016-0609>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	141
F/Sty/Disfl	86
A/MisTra/SemUn	73
F/Lex/LexCho	41
A/Omis	25
G/Orth/Capit	19
A/MisTra/Part	18
A/Untra	12
A/MisTra/Gend	11
F/SLP/Nomin	11
A/MisTra/WSD	8
A/MisOrd/MW	7
A/MisTra/WroPr	6
G/WF/Agree	6

F/Sty/Reg	5
F/De/Layo	4
A/Mech/Capit	4
A/MisOrd/WO	4
A/MisTra/POS	4
G/Orth/Punct	4
A/Add	3
A/ForLa	3
A/Termi/SemUn	3
F/Coh/CoRef	3
F/Coh/Cit	2
G/Orth/WhiSpa	2
G/Orth/Spel	1
A/MisExpl	1
A/SouEr	1
A/Termi/DNT	1
A/MWE/Ell	1
F/SLP/ForImp	1

Tabella 13 (All. 2): Analisi quantitativa degli errori per il testo n. 7 del progetto n. 1

In fase di individuazione degli errori, ne vengono riscontrati 294 di *Fluency*, 185 di *Accuracy* e 32 di *Grammar*. Il totale degli errori annotati ammonta a 511. La maggior parte degli errori appartiene alla macrocategoria *Fluency*, soprattutto per l'alta frequenza della sottocategoria F/Coh/InPu (141 occorrenze) che segnala l'uso incoerente di virgolette e apostrofi. Rilevanti sono anche le seguenti categorie: F/Sty/Disfl per gli errori di scorrevolezza della lingua di arrivo (86), A/MisTra/SemUn per gli errori di mancata corrispondenza semantica (73) e F/Lex/LexCho per gli errori di scelta lessicale in riferimento all'ambito e al genere testuale trattato (41). Si registra un numero significativo di errori di segmenti del testo di partenza omessi nel testo di arrivo (A/Omis, 25). In merito gli errori stilistici del linguaggio scientifico, se ne constatano 11 di nominalizzazione mancata (F/SLP/Nomin), 5 di registro inadeguato (F/Sty/Reg) e un errore di mancato uso della forma impersonale. Gli errori terminologici risultano invece limitati, data la scarsa presenza di terminologia specialistica, con 3 occorrenze per la categoria A/Termi/SemUn (traduzione errata) e una per A/Termi/DNT (traduzione non necessaria). Infine, sono stati annotati 2 errori di F/Coh/Cit, riguardanti la strutturazione incoerente dei riferimenti bibliografici.

Durata delle sessioni di post-editing con Qualityity

Il *post-editing* si è articolato in cinque sessioni della durata rispettiva di 04:28:56, 02:30:42, 02:01:19, 00:58:26 e 00:58:32. Il totale delle sessioni ammonta a 10:57:55.

Report di Post-Edit Compare

Gli interventi di *post-editing* vengono effettuati su un testo che comprende 215 segmenti (per un totale di circa 3.945 parole). La quasi totalità dei segmenti, 211, è stata generata dal sistema di TA, mentre solo 4 segmenti risultano immessi manualmente come *New*. Anche in questo caso non si registrano contributi dalla memoria di traduzione, a conferma che la base di partenza è interamente costituita dall'output della TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	49	480	12,17%
95-99%	29	539	13,66%
85-94%	59	1.497	37,95%
75-84%	22	499	12,65%
50-74%	21	475	12,04%
New (0-49%)	36	455	11,53%

Tabella 14 (All. 2): Distribuzione delle fasce di PEMp per testo n.7 del progetto n. 1

L'analisi delle fasce di PEMp mostra una distribuzione equilibrata, ma con alcuni elementi di interesse specifico. La fascia 100% raccoglie il 12,17% delle parole, segnalando che una porzione molto limitata del testo prodotto dalla macchina è stata accettata integralmente senza alcuna correzione. Si tratta di un dato minoritario che suggerisce una affidabilità del sistema di TA per porzioni limitate del testo. Gli interventi più significativi si collocano nella fascia 85-94%, che riguarda il 37,95% delle parole. Questo valore indica che più di un terzo del testo è stato interessato da correzioni di lieve entità. La fascia 95-99% raccoglie il 13,66% e indica interventi minimi ma non trascurabili. Limitati sono anche i valori per la fascia 75-84% con il 12,65% e la fascia 50-74% con il 12,04%, dove gli interventi comportano modifiche più consistenti. Infine, la categoria *New* (0-49%) copre l'11,53% delle parole, a indicare segmenti interessati da rielaborazioni largamente consistenti fino a integrali.

Complessivamente, si può affermare che l'output fornisce una base di partenza utile, ma non priva di criticità: solo circa un decimo del testo è stato accettato senza variazioni, mentre quasi il 90% ha richiesto un vario grado di intervento umano, una prevalenza di correzioni lievi. La distribuzione disomogenea delle fasce di intervento suggerisce che, pur essendo la TA sufficientemente vicina a una resa accettabile, il lavoro del *post-editor* resta determinante per garantire qualità e coerenza.

Testo n. 8 (8.744 tokens)

Niewalda, Katrin, Maria Gabriela Schmidt, e Shinichi Sakamoto. 2016. «Ergänzungsvorschläge für Hörverstehensübungen in deutschen Lehrwerken im universitären DaF-Unterricht in Japan».

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	177
F/Sty/Disfl	111
A/MisTra/SemUn	99
F/Lex/LexCho	99
A/Termi/SemUn	37
A/Omis	30
A/ForLa	27
G/Orth/Capit	20
A/DNT	19
G/Orth/Punct	19
G/Orth/WhiSpa	17
F/Sty/Reg	16
A/MisTra/WSD	15
A/Untra	15
A/MisTra/Part	14
F/SLP/Nomin	12
A/MisTra/WroPr	12
A/MisTra/Gend	12
A/MisOrd/WO	9
A/MisExpl	7
F/Lex/Fue	6
G/Orth/Spel	5
A/Mech/Punct	5
A/MisTra/POS	5
G/WF/Agree	4
A/MisOrd/MW	4
A/MWE/Komp	2
F/Coh/CoRef	2
A/Termi/DNT	2
A/Add	1
F/Sty/Idiom	1
A/MWE/Ell	1
F/De/Layo	1
F/De/CharFor	1

Tabella 15 (All. 2): Analisi quantitativa degli errori per il testo n. 8 del progetto n. 1

Nell'articolo scientifico esaminato, si annotano complessivamente 807 errori, suddivisi in 426 di *Fluency*, 316 di *Accuracy* e 65 di *Grammar*. Gli errori di *Fluency* superano la metà del

totale, soprattutto a causa della frequenza molto elevata della categoria F/Coh/InPu, che conta 177 occorrenze e segnala i casi di uso incoerente delle virgolette e dell’apostrofo. Seguono, per numerosità, le seguenti categorie: F/Sty/Disfl per le problematiche di scorrevolezza della lingua di arrivo (111); A/MisTra/SemUn per gli errori di mancata corrispondenza semantica tra segmenti di arrivo e segmenti di partenza (99) e F/Lex/LexCho per la scelta di parole inadeguate all’ambito e al genere testuale trattato (99). Si rileva anche l’omissione di segmenti del testo di partenza nel testo di arrivo (A/Omis, 30 occorrenze) e l’uso improprio della lingua inglese o di parole inesistenti nella lingua di arrivo (A/ForLa, 27 occorrenze). Con riferimento agli aspetti stilistici legati al linguaggio scientifico, si rilevano 16 errori di registro inadeguato (F/Sty/Reg), 12 errori di nominalizzazione non impiegata (F/SLP/Nomin), 6 errori di traduzione delle *Fügungen* (F/Lex/Fue). Sul piano terminologico si registrano invece 37 errori di traduzione errata (traduzione errata) e 2 errori di traduzione non necessaria (A/Termi/DNT).

Durata delle sessioni di post-editing con Qualityity

Il processo di *post-editing* è stato articolato in sette sessioni della durata rispettiva di 01:09:21, 03:44:14, 02:19:35, 01:53:42, 01:56:56, 01:16:12 e 04:33:54, per un totale complessivo di 16:53:54.

Report di Post-Edit Compare

Il report attinente a questo testo comprende 374 segmenti (per un totale di circa 7.915 parole). La quasi totalità, 367 segmenti, è generata dal sistema di TA DeepL, mentre solo 7 segmenti risultano introdotti come *New*. Anche in questo caso non si osservano *matches* riconducibili all’impiego dalla memoria di traduzione nel progetto di traduzione, confermando che il testo di partenza analizzato proviene interamente dal sistema di TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	86	1.269	18,35%
95-99%	34	760	10,99%
85-94%	95	2.258	32,65%
75-84%	43	888	12,84%
50-74%	38	611	8,84%
New (0-49%)	78	1.129	16,33%

Tabella 16 (All. 2): Distribuzione delle fasce di PEMp per testo n. 8 del progetto n. 1

La distribuzione delle modifiche, riportata sottoforma di fasce percentuali nella Tabella 31, evidenzia un quadro interessante tra segmenti accettati, correzioni parziali e riscritture più ampie. La fascia 100% raccoglie il 18,35% delle parole, evidenziando che quasi un quinto del testo è stato riutilizzato senza alcuna modifica: un dato significativo, che mostra una buona resa del sistema in

porzioni non trascurabili del documento. La parte significativa delle modifiche si concentra nella fascia 85-94%, che rappresenta il 32,65%, indicando che circa un terzo del testo è stato sottoposto a interventi di media entità. Seguono la fascia 95-99% con il 10,99% riguardante modifiche minime e mirate, e la fascia 75-84% con il 12,84%, che evidenzia interventi di media intensità. La fascia 50-74% copre l'8,84%, indice di modifiche più ampie su una porzione limitata di segmenti. Infine, la categoria *New* (0-49%) rappresenta il 16,33% delle parole, mostrando che oltre un sesto del testo è stato riscritto integralmente o quasi.

Complessivamente, emerge un output della TA che fornisce una base non omogenea: una quota consistente di segmenti è stata mantenuta invariata, mentre altre parti hanno richiesto correzioni di varia entità, privilegiando in particolar modo le modifiche di lieve entità. Al risultato consistente per la fascia 100% si contrappone un risultato altrettanto consistente per la fascia *New* (0-49%), suggerendo pertanto una qualità discontinua, con risultati ottimali in alcuni segmenti e del tutto insoddisfacenti in altri. In questo scenario, l'intervento del *post-editor* risulta indispensabile.

Testo n. 9 (10.187 tokens)

Wichmann, Martin. 2016. «Jugendsprache in DaF-Lehrwerken – oft gefordert, aber selten praktiziert. Eine Bestandsaufnahme». *Informationen Deutsch als Fremdsprache* 43 (6): 667–92. <https://doi.org/doi:10.1515/infodaf-2016-0608>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	244
F/Sty/Disfl	132
F/Lex/LexCho	121
A/MisTra/SemUn	99
A/ForLa	80
A/DNT	76
G/Orth/Capit	41
A/MisTra/WSD	33
A/Omis	31
A/Untra	25
A/Termi/SemUn	24
F/SLP/ForImp	21
G/Orth/WhiSpa	19
F/SLP/Nomin	18
A/MisExpl	18
F/Sty/Reg	17
A/MisOrd/WO	12
A/MisTra/Part	12

F/Lex/Fue	11
A/MisTra/POS	10
F/Coh/Cit	8
A/SouEr	6
A/MisTra/WroPr	5
G/Orth/Punct	5
A/Mech/Punct	4
A/Add	3
A/MisOrd/MW	2
F/Sty/Idiom	2
G/Orth/Spel	2
F/Coh/CoRef	1
F/De/Layo	1
G/WF/Agree	1
G/WF/VF	1

Tabella 17 (All. 2): Analisi quantitativa degli errori per il testo n. 9 del progetto n. 1

Nel testo si riscontrano 576 errori di *Fluency*, 440 di *Accuracy* e 69 di *Grammar*. Il totale degli errori annotati ammonta a 1.085 errore. La quota più ingente riguarda la macrocategoria *Fluency*, dovuta in particolare alla frequenza degli errori riguardanti l'uso incoerente di virgolette e apostrofo (F/Coh/InPu, 244 occorrenze), la scorrevolezza della lingua di arrivo (F/Sty/Disfl, 132 occorrenza) e la scelta lessicale (F/Lex/LexCho, 121 occorrenze). Rilevanti sono anche i seguenti errori di *Accuracy*: A/MisTra/SemUn per l'assenza di corrispondenza semantica tra il segmento di arrivo e quello di partenza (99 occorrenze), A/ForLa per il mancato uso di parole nella lingua di arrivo (80 occorrenze) e A/DNT per i segmenti da non tradurre (76 occorrenze). L'elevato numero di occorrenze per la categoria A/ForLa è riconducibile alla frequente traduzione dell'acronimo tedesco DaF (*Deutsch als Fremdsprache*) con l'acronimo inglese GFL (*German as a Foreign Language*), non usato comunemente in lingua italiana. Gli errori di A/DNT derivano invece dalla presenza di esempi in lingua tedesca, la cui traduzione in lingua italiana non si ritiene necessaria. Si registrano inoltre 41 errori per la categoria grammaticale che segnala un uso errato della lettera maiuscola (G/Orth/Capit); 33 errori semantici per parole polisemiche (A/MisTra/WSD) e 31 errori di omissione di segmenti del testo di partenza nel testo di arrivo (A/Omis). Nell'ambito stilistico legato al linguaggio scientifico, vengono riscontrati 21 errori di mancato uso della forma impersonale (F/SLP/ForImp), 18 errori di nominalizzazione non impiegata (F/SLP/Nomin), 17 errori di registro inadeguato (F/Sty/Reg) e 11 errori di traduzione di *Fügungen* (F/Lex/Fue). In riferimento alla componente terminologica si osservano 24 errori di traduzione (A/Termi/SemUn). Gli errori di F/Coh/Cit, relativi alla strutturazione incoerente dei riferimenti bibliografici, ammontano a 24.

Durata delle sessioni di post-editing con Quality

Il processo di *post-editing* è stato condotto in sei sessioni, della durata rispettiva di 02:20:37, 01:58:26, 00:49:15, 01:00:21, 04:24:04 e 04:26:59, per un totale complessivo di 14:59:42.

Report di Post-Edit Compare

Il *post-editing* ha riguardato 504 segmenti, pari a circa 8.157 parole. La quasi totalità del testo, 495 segmenti, è stata tradotta automaticamente, mentre solo 9 segmenti sono introdotti come *New*. La memoria di traduzione non ha contribuito fornendo matches, pertanto si può dedurre che la base di partenza fosse interamente costituita dall'output elaborato dal sistema di TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	149	1.689	20,71%
95-99%	47	893	10,95%
85-94%	90	1.950	23,91%
75-84%	62	1.300	15,94%
50-74%	57	973	11,93%
New (0-49%)	99	1.352	16,57%

Tabella 18 (All. 2): Distribuzione delle fasce di PEMp per testo n. 9 del progetto n. 1

L'analisi della distribuzione delle modifiche mostra un quadro variegato. La fascia 100% raccoglie il 20,71% delle parole, il valore più alto raggiunto finora, che pone in evidenza che un quinto del testo è stato accettato senza alcuna correzione. Questo dato mette in luce la capacità della macchina di produrre, in misura non trascurabile, segmenti di qualità pubblicabile. La fascia 95-99%, con il 10,95% delle parole, mostra che i micro-interventi trovano una diffusione limitata in questo testo. Diversamente, la fascia 85-94% presenta il valore più elevato, il 23,91% indicando che quasi un quarto del testo ha subito delle modifiche lievi. Le fasce che rappresentano interventi più sostanziali registrano i seguenti valori: 15,94% per la fascia 75-84% e il 11,93% per quella 50-74%. Qui il lavoro di *post-editing* è stato più incisivo, anche se su una porzione limitata, ma non trascurabile, di testo. Infine, la categoria *New* (0-49%) comprende il 16,57% delle parole, indice di riscritture estremamente ampie e/o complete per circa un sesto del testo.

Dall'analisi del report emerge quindi un output disomogeneo: da un lato un buon numero di segmenti (oltre il 20%) accettati integralmente, dall'altro una quota altrettanto significativa di segmenti interessati da rielaborazioni consistenti fino a integrali. Questa polarizzazione evidenzia come il sistema di TA alterni risultati di alta qualità a traduzioni insoddisfacenti, richiedendo interventi che oscillano tra modifiche in ampia misura lievi fino a correzioni più o meno radicali.

Riflessione conclusiva

Complessivamente le tabelle quantitative dei singoli testi mostrano che la macrocategoria di errori *Fluency* presenta il numero più elevato di occorrenze. Fa eccezione il testo n. 4, nel quale la macrocategoria *Accuracy* risulta predominante. Nella maggior parte dei casi, alla *Fluency* seguono la macrocategoria *Accuracy* e *Grammar*; quest'ultima registra costantemente il numero più contenuto di occorrenze. I dati precedentemente illustrati suggeriscono che, sotto il profilo quantitativo, le criticità più rilevanti riguardano la scorrevolezza e la naturalezza della lingua di arrivo, che si conferma un fattore decisivo per garantire la qualità pubblicabile del testo scientifico. Per un'analisi qualitativa approfondita delle diverse tipologie di errore si rimanda al capitolo 6, paragrafo 6.1.

L'analisi congiunta del numero di *tokens* e della durata complessiva del *post-editing* mostra una tendenza generale: i testi più estesi richiedono in media un tempo di *post-editing* superiore. Tuttavia, tale rapporto non risulta costantemente lineare. Ad esempio, il testo n. 8, pur presentando un numero di *tokens* inferiore rispetto al testo n. 9 (rispettivamente 8.744 e 10.187), richiede una durata di *post-editing* maggiore (16:53:54 contro 14:59:42). Questo dato suggerisce che l'estensione del testo non costituisce di per sé un indicatore sufficiente a determinare il *post-editing effort*, poiché intervengono ulteriori fattori quali la complessità sintattica e terminologica, la densità informativa e, soprattutto, la qualità dell'output prodotto dal sistema di TA.

Un andamento analogo emerge dal confronto tra il numero di errori annotati e la durata del *post-editing*. Anche in questo caso, sebbene si osservi una relazione di proporzionalità generale secondo cui i testi con un numero più elevato di errori richiedono tempi di correzione più consistenti, non mancano eccezioni significative. Il testo n. 5, ad esempio, presenta un numero di errori complessivamente inferiore rispetto al testo n. 6 (rispettivamente 575 e 785), tuttavia, i tempi di *post-editing* sono pressoché equivalenti (rispettivamente 12:57:45 e 13:34:09). Ciò conferma che non è tanto la quantità di errori a incidere sui tempi, quanto piuttosto la loro natura e complessità.

Infine, considerando il rapporto tra il numero di *tokens* e i tempi di correzione, è possibile stimare una produttività media compresa tra circa 500 e 700 *tokens* l'ora. Questa variabilità riflette la qualità eterogenea dell'output tradotto e la conseguente necessità di interventi umani di varia entità.

Per quanto riguarda i report di Post-Edit Compare, si può affermare che la fascia 85-94% (correzioni lievi) rappresenta la quota più consistente, con valori compresi tra il 24% e il 38%, fatta eccezione per il testo n. 6, in cui raggiunge circa il 19%, risultando il secondo valore più alto. Per la categoria *New* (0-49%), che segnala correzioni totali o quasi integrali del segmento, si osservano due tendenze: una prima con valori compresi tra il 28% e il 19% (cinque testi), e una seconda con

valori tra il 16-11% (quattro testi). Anche nella fascia 100% (segmenti non modificati) emergono due gruppi distinti: tre testi presentano valori tra il 21-16% e sei tra il 9-13%. Le fasce 95-99% (correzioni minime), 75-84% (correzioni medie) e 50-74% (correzioni medio-alte) presentano valori generalmente compresi tra l'8% e il 18%, fatta eccezione per la fascia 75-84% nel testo n. 4, che raggiunge circa il 23%. Sull'intero progetto di traduzione si evidenziano quindi tendenze specifiche a livello di singoli testi, piuttosto che pattern uniformi. In particolare, i valori relativamente contenuti nella fascia 100% confermano che il sistema di TA produce raramente segmenti utilizzabili senza modifiche, richiedendo invece interventi di riformulazione per adeguare il testo ai parametri di coerenza e leggibilità richiesti nella scrittura scientifica.

Allegato 3 – Progetto n. 2 *trans-kom*: analisi dei singoli testi

L'analisi si articola in più parti: una tabella degli errori annotati con un commento alle categorie principali, le tempistiche delle sessioni di *post-editing* e, infine, una presentazione dei risultati del report di Post-Edit Compare mediante tabelle con relativo commento.

Testo n.1 (12.749 tokens)

Kempa, Thomas. 2013. «Anregungen für eine Differenzierung eines funktionalen translationsdidaktischen Verstehens- und Sinnbegriffs aus pragmatistischer Sicht». *trans-kom* 6 (2): 345–70.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	297
F/Sty/Disfl	105
A/MisTra/SemUn	97
A/MisTra/WSD	79
F/Lex/LexCho	76
F/Sty/Reg	55
A/Omis	45
A/Termi/SemUn	43
A/SouEr	27
G/Orth/WhiSpa	26
G/Orth/Punct	23
A/DNT	20
F/SLP/ForImp	19
F/Lex/Fue	19
A/MisTra/Part	13
F/SLP/Nomin	12
G/Orth/Spel	12
A/Untra	11
A/MWE/Komp	11
A/MisOrd/WO	9
G/WF/Agree	9
G/Orth/Capit	8
A/Mech/CharFor	8
A/Mech/Punct	7
A/MisTra/POS	6
A/Add	6
F/Coh/CoRef	6
A/MisTra/WroPr	5
A/MisTra/Gend	4

F/Sty/Idiom	3
F/Coh/Cit	3
G/WF/VF	3
A/MWE/Ell	2
A/Termi/DNT	2
F/De/Layo	2
A/MisExpl	1
F/De/CharFor	1

Tabella 1 (All. 3): Analisi quantitativa degli errori per il testo n. 1 del progetto n. 2

Nel testo preso in esame si individuano complessivamente 1.075 errori, suddivisi in 598 di *Fluency*, 396 di *Accuracy* e 81 di *Grammar*. La maggior parte degli errori appartiene alla macrocategoria *Fluency*, in cui la sottocategoria F/Coh/InPu riveste un ruolo fondamentale registrando 297 occorrenze. Questi errori riguardano l'uso incoerente delle virgolette e dell'apostrofo. Seguono, per numero di occorrenze, la categoria F/Sty/Disfl per le anomalie di scorrevolezza nella lingua di arrivo (105 occorrenze), A/MisTra/SemUn per la mancata corrispondenza semantica tra testo di partenza e testo di arrivo (97 occorrenze) e A/MisTra/WSD per la scelta semantica errata in riferimento a parole polisemiche (79 occorrenze). Per quanto riguarda gli errori stilistici riguardanti il linguaggio scientifico, si constatano 76 errori di lessico inadeguato all'ambito e al genere testuale (F/Lex/LexCho), 55 errori di registro inappropriato (F/Sty/Reg), 19 errori di traduzione di *Fügungen* (F/Lex/Fue), 19 errori di forma impersonale non impiegata (F/SLP/ForImp) e 12 errori di nominalizzazione mancata (F/SLP/Nomin). La componente terminologica presenta 43 errori di traduzione di termini (A/Termi/SemUn) e solo 2 di traduzione non necessaria di termini (A/Termi/DNT). Altre categorie con frequenza significativa includono A/Omis (45 errori) e A/SouEr (27 errori). Gli errori di A/Omis riguardano l'omissione di segmenti, parole o frasi intere, nel testo di arrivo. Diversamente, gli errori di A/SouEr segnalano le problematiche presenti nel testo di partenza e in questo caso derivano dal posizionamento errato del numero della nota a piè di pagina, errore che si riscontra nel file traducibile generato da Trados Studio, e quindi anche nel testo di arrivo, ma non nel testo originale. La categoria F/Coh/Cit, rilevante per il genere testuale dell'articolo scientifico, registra solo 3 occorrenze, indicando una scarsa presenza di errori legati a una strutturazione incoerente delle citazioni.

Durata delle sessioni di post-editing con Qualityity

Il *post-editing* è stato condotto in otto sessioni della durata rispettiva di 02:17:43, 01:01:02, 01:08:26, 01:55:45, 04:28:26, 02:15:13, 02:00:14 e 03:11:01, per un totale complessivo di 18:17:50.

Report di Post-Edit Compare

Nella fase di *post-editing* si sono presi in esame 405 segmenti (per un totale di circa 11.307 parole). 369 segmenti del totale vengono generati dal sistema di TA, mentre 36 segmenti risultano introdotti come “nuova” traduzione. Non si rilevano proposte formulate dalla memoria di traduzione, confermando che l’output analizzato proviene interamente dal sistema di TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	56	967	9,38%
95-99%	73	2.144	20,80%
85-94%	127	3.679	35,69%
75-84%	53	1.453	14,10%
50-74%	29	722	7%
New (0-49%)	67	1.342	13,02%

Tabella 2 (All. 3): Distribuzione delle fasce di PEMp per testo n. 1 del progetto n. 2

La tabella 35 mostra una distribuzione variegata tra segmenti accettati e modifiche di diversa entità. La fascia 100% raccoglie il 9,38% delle parole, indicando che quasi un decimo del testo è stato riutilizzato integralmente senza modifiche e suggerendo che una parte molto limitata è stata tradotta adeguatamente da DeepL. La fascia 95-99% rappresenta il 20,80% delle parole, comprendendo interventi minimi per circa un quarto del testo. La fascia 85-94% copre la quota più ampia, con il 35,69% delle parole, evidenziando oltre un terzo del testo è stato interessato da correzioni di lieve entità. Interventi di *post-editing* più consistenti sono rappresentati dalle fasce 75-84% con il 14,10% e 50-74% con il 7%, mettendo in luce una distribuzione limitata di questi tipi di correzioni. Infine, la categoria *New* (0-49%) copre il 13,02% delle parole (67 segmenti, 1.342 parole), indicando una porzione limitata ma non trascurabile di segmenti riscritti in larga misura o totalmente.

Complessivamente, l’output della TA fornisce una base di partenza utile ma non uniforme: meno di un decimo del testo è stato accettato senza interventi, mentre circa il 90% richiede revisioni di diversa entità. Dal punto di vista dell’*effort*, le operazioni di *post-editing* includono in gran parte interventi di minima e lievi entità, non escludendo tuttavia interventi più approfonditi e riscritture integrali.

Testo n. 2 (7.121 tokens)

Pavlova, Anna. 2014. «Strategie der Übersetzung und Beurteilung der Übersetzungsqualität». *trans-kom* 7 (2): 256–71.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	159
A/Untra	61
F/Sty/Disfl	56
A/MisTra/SemUn	53
F/Lex/LexCho	51
G/Orth/Capit	31
A/Omis	21
A/MisTra/WSD	18
A/Termi/SemUn	18
A/DNT	18
F/Sty/Reg	15
A/MisTra/Part	11
F/SLP/ForImp	9
A/MisExpl	9
G/Orth/Punct	8
F/SLP/Nomin	8
A/ForLa	7
G/Orth/WhiSpa	6
A/MisOrd/WO	5
G/WF/Agree	5
F/Coh/CoRef	5
A/SouEr	4
A/Mech/Punct	4
A/MisTra/POS	4
F/Sty/Idiom	4
A/MisTra/WroPr	3
F/Lex/Fue	2
G/Orth/Spel	2
A/Mech/CharFor	2
G/WF/VF	2
F/De/Layo	2
F/Coh/Cit	2

Tabella 3 (All. 3): Analisi quantitativa degli errori per il testo n. 2 del progetto n. 2

In fase di analisi del testo si rilevano 313 errori di *Fluency*, 238 errori di *Accuracy* e 54 errori di *Grammar*. Il totale degli errori annotati ammonta a 605. La maggior parte degli errori appartiene alla categoria *Fluency*, in particolare F/Coh/InPu che segnala 159 casi di uso incoerente delle virgolette e dell’apostrofo. Seguono, per numero di occorrenze, le seguenti categorie: A/Untra per i segmenti non tradotti (63 errori); F/Sty/Disfl per le problematiche di scorrevolezza della lingua di arrivo (56 errori); A/MisTra/SemUn per gli errori di corrispondenza semantica tra il testo di partenza e quello di arrivo (53 errori) e F/Lex/LexCho per la scelta lessicale errata in riferimento all’ambito e genere testuale trattati (51 errori). Rilevanti sono anche gli errori di uso della lettera

maiuscola (G/Orth/Capit) individuati in 31 casi nel testo. Nell'ambito degli errori stilistici relativi al linguaggio scientifico, si segnalano 15 errori di registro inadeguato (F/Sty/Reg), 9 errori di forma impersonale non impiegata (F/SLP/ForImp), 8 errori di nominalizzazione mancata (F/SLP/Nomin), 2 errori di traduzione di *Fügungen* (F/Lex/Fue). La componente terminologica presenta 18 occorrenze per la categoria A/Termi/SemUn che segnala le traduzioni errate di termini. La categoria F/Coh/Cit registra 2 errori, indicando solo due casi di strutturazione incoerente dei riferimenti bibliografici.

Durata delle sessioni di post-editing con Qualityity

Il lavoro di *post-editing* è stato articolato in cinque sessioni, della durata rispettiva di 00:44:03, 01:16:52, 04:02:42, 03:23:21 e 00:57:20, per un totale complessivo di 10:24:18.

Report di Post-Edit Compare

Il report riguardante questo testo comprende 337 segmenti (per un totale di circa 5.763 parole). Quasi la totalità dei segmenti, 328, è stata generata dal sistema di TA, mentre 7 segmenti risultano introdotti come *New*. Non si registrano contributi della memoria di traduzione (*matches* di vario tipo), pertanto che l'output prese in esame ha origine interamente dal sistema di TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	90	976	16,94%
95-99%	46	1.138	19,75%
85-94%	66	1.554	26,97%
75-84%	39	847	14,70%
50-74%	35	453	7,86%
New (0-49%)	61	795	13,79%

Tabella 4 (All. 3): Distribuzione delle fasce di PEMp per testo n. 2 del progetto n. 2

La Tabella 37 evidenzia una distribuzione interessante della fasce di PEMp. La fascia 100% raccoglie il 16,94% delle parole, suggerendo che quasi un sesto del testo è stato accettato senza modifiche, indice di una buona resa del sistema su determinate porzioni del documento. La fascia 95-99%, che segnala interventi minimi, registra il 19,75% interessando quasi un quinto dell'intero testo. Invece, la fascia 85-94% copre il 26,97% delle parole, indicando che oltre un quarto del testo è stato sottoposto a interventi di lieve entità. Seguono la fascia 75-84%, che presenta il 14,70% e la fascia 50-74% con il 7,86%. I dati delle due ultime categorie, piuttosto limitati, riguardano correzioni di media e medio-alta intensità. Infine, la categoria *New* (0-49%), che riguarda segmenti riscritti in ampia misura o integralmente, copre il 13,79% delle parole.

Nel complesso, l'analisi evidenzia una distribuzione equilibrata tra segmenti accettati integralmente, correzioni minime, medie e riscritture più consistenti. Una porzione limitata del testo presentava una qualità pubblicabile; la restante parte è stata sottoposta a modifiche di varia entità, tra le quali spiccano gli interventi di lieve intensità. In sintesi, l'output della TA è da considerarsi funzionale come base iniziale, ma non omogeneo nella qualità dei segmenti, facendo emergere come il lavoro del *post-editor* sia fondamentale per garantire qualità, coerenza e leggibilità complessiva.

Testo n. 3 (9.841 tokens)

Hilgert, Mariana. 2015. «Übersetzung zwischen möglichen Welten – eine Suche nach einem theoretischen Ausweg». *trans-kom* 8 (2): 539–59.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	249
A/MisTra/SemUn	75
F/Sty/Disfl	58
F/Lex/LexCho	27
F/De/Layo	22
G/Orth/Capit	19
G/Orth/Punct	19
A/DNT	19
A/Mech/CharFor	18
A/Omis	18
F/Sty/Reg	16
F/SLP/ForImp	15
A/MisTra/Part	12
A/Mech/Capit	12
G/Orth/WhiSpa	11
A/Untra	9
G/WF/Agree	8
A/ForLa	8
A/Termi/SemUn	7
A/MisTra/WSD	6
A/SouEr	6
F/SLP/Nomin	6
A/MisOrd/WO	5
A/MisTra/WroPr	5
F/Lex/Fue	4
G/Orth/Spel	4

F/Sty/Idiom	3
A/Add	3
A/MisTra/POS	3
A/Oth	2
F/Coh/CoRef	2
A/MisExpl	1
G/WF/VF	1

Tabella 5 (All. 3): Analisi quantitativa degli errori per il testo n. 3 del progetto n. 2

Nel testo vengono identificati 402 errori di *Fluency*, 209 errori di *Accuracy* e 62 errori di *Grammar*, per un totale di 673 errori annotati. La maggior parte di questi errori riguarda la macrocategoria *Fluency*, con particolare incidenza della categoria F/Coh/InPu, che presenta la quota più consistente di errori con 249 occorrenze. Questa categoria segnala l'uso incoerente delle virgolette e dell'apostrofo. Rilevanti, seppur in misura minore, sono anche gli errori riguardanti la mancata corrispondenza semantica tra il testo di arrivo e il testo di partenza (A/MisTra/SemUn, 75 occorrenze) e gli errori di scorrevolezza della lingua di arrivo (F/Sty/Disfl, 58 occorrenze). Si constatano anche 22 errori di F/De/Lay relativi alla strutturazione del testo e dei paragrafi. Gli errori stilistici riferibili al linguaggio scientifico si distribuiscono come segue: 27 errori di lessico in riferimento all'ambito e al genere testuale trattati (F/Lex/LexCho), 16 errori di registro inadeguato (F/Sty/Reg), 15 errori di forma impersonale non impiegata (F/SLP/ForImp), 6 errori di nominalizzazione non realizzata (F/SLP/Nomin) e 4 errori di traduzione di *Fügungen* (F/Lex/Fue). La componente terminologica si rivela invece meno problematica, con solo 7 errori di A/Termi/SemUn che segnalano la traduzione errata di termini.

Durata delle sessioni di post-editing con Qualitivity

La fase di *post-editing* è stata articolata in cinque sessioni, della durata rispettiva di 03:28:55, 01:03:17, 02:16:12, 02:40:59 e 01:18:32, per un totale complessivo di 10:47:55.

Report di Post-Edit Compare

Il report di Post-Edit Compare relativo a questo documento comprende 371 segmenti, per un totale di circa 7.978 parole. La maggior parte, 334 segmenti, è stata generata dal sistema di TA, mentre 37 segmenti risultano inseriti come *New*. La memoria di traduzione non ha contribuito al testo proponendo dei *matches*, il che conferma che il testo di partenza dell'analisi è stato quasi interamente prodotto dalla TA.

Fasce di PEMP	Segmenti	Parole	Percentuali
100%	87	1.207	15,13%
95-99%	81	2.255	28,27%

85-94%	66	1.908	23,92%
75-84%	25	728	9,13%
50-74%	35	709	8,89%
New (0-49%)	79	1.171	14,68%

Tabella 6 (All. 3): Distribuzione delle fasce di PEMp per testo n. 3 del progetto n. 2

La tabella 39 illustra le fasce di PEMp con le rispettive percentuali evidenziando una distribuzione interessante. La fascia 100% raccoglie il 15,13% delle parole, cioè i casi in cui l'output è stato accettato integralmente, senza interventi. Si tratta di un dato significativo, che indica una discreta capacità del sistema di produrre traduzioni già utilizzabili. La fascia 95-99%, che presenta il 28,27% delle parole, è la fascia più rilevante: qui le correzioni richieste sono minime e mirate, suggerendo che oltre un quarto del testo era già vicino a una resa accettabile. Consistente è anche la quota per la fascia 85-94%, che con il 23,92% copre quasi un quarto del testo. La fascia 75-84% con il 9,13% e quella 50-74% con l'8,89% segnalano una diffusione limitata di modifiche di entità media e medio-alta. Infine, la categoria *New* (0-49%) raggiunge il 14,68%, indicando un volume limitato ma non trascurabile di riscritture integrali o semi-integrali.

Nel complesso, il profilo di questo documento appare polarizzato: da un lato una quota limitata di segmenti resta invariata, dall'altro il 52% del testo ha richiesto modifiche di entità minima o lieve (fasce 95-99% e 85-945) e una quota non trascurabile riscritture parziali o totali (quasi il 25% tra 50-74% e *New*). Questa combinazione rivela un output eterogeneo. Dal punto di vista del *post-editing*, ciò si traduce in un *effort* disomogeneo: nel *post-editing* si alternano interventi di varia entità, da modifiche minime fino a riscritture radicali.

Testo n. 4 (5.701 tokens)

Schum, Swenia. 2017. «Barrierefreiheit als Herausforderung in der Fachtextübersetzung». *transkom* 10 (3): 349–63.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Sty/Disfl	70
F/Coh/InPu	69
A/Termi/SemUn	65
F/Lex/LexCho	60
A/MisTra/SemUn	49
F/De/CharFor	49
A/Termi/DNT	34
G/Orth/Capit	27
A/Omis	16
A/Untra	16

G/Orth/WhiSpa	11
A/MisTra/WSD	10
F/Sty/Reg	9
A/MisOrd/WO	9
F/SLP/Nomin	8
A/MisTra/Part	5
G/Orth/Punct	4
A/MisExpl	4
A/SouEr	3
F/SLP/ForImp	3
F/Coh/Cit	3
G/Orth/Spel	3
A/Mech/Punct	3
A/MisTra/POS	3
A/DNT	2
F/Lex/Fue	2
A/Mech/CharFor	2
A/Add	2
A/MisTra/WroPr	2
A/MWE/Ell	2
F/De/Layo	2
A/ForLa	2
F/Coh/CoRef	2

Tabella 7 (All. 3): Analisi quantitativa degli errori per il testo n. 4 del progetto n. 2

Si rilevano complessivamente 551 errori, di cui 277 di *Fluency*, 229 di *Accuracy* e 45 di *Grammar*. Gli errori di *Fluency* sono la maggioranza del totale degli errori. Le seguenti categorie sono le più rappresentate (in ordine decrescente di occorrenze): F/Sty/Disfl per le problematiche di scorrevolezza della lingua di arrivo (70 occorrenze); F/Coh/InPu per l'uso incoerente di virgolette e apostrofo (69 occorrenze); A/Termi/SemUn per la traduzione errata di termini (65 occorrenze); F/Lex/LexCho per la scelta lessicale inadeguata all'ambito e al genere testuale trattati (60 occorrenze); A/MisTra/SemUn per la mancata corrispondenza semantica tra segmenti di arrivo e di partenza e F/De/CharFor per la formattazione errata secondo le convenzioni linguistiche del testo di arrivo (49 occorrenze). Per quanto concerne la componente terminologica, oltre alla forte presenza di errori in A/Termi/SemUn, vengono annotati 34 casi di A/Termi/DNT, legati alla traduzione non necessaria di un termine. Questi errori sono associati sempre all'errore F/De/CharFor, dovuto alla necessità di riportare in corsivo i termini stranieri nel testo di arrivo. Dal punto di vista stilistico del linguaggio scientifico, si constata 9 errori di registro inadeguato (F/Sty/Reg), 8 errori di nominalizzazione non realizzata (F/SLP/Nomin), e 3 errori di forma impersonale non impiegata

(F/SLP/ForImp) e 2 errori di traduzioni di *Fügungen* (F/Lex/Fue). Gli errori classificati come F/Coh/Cit (3 occorrenze) riguardano la composizione non coerente delle citazioni bibliografiche.

Durata delle sessioni di post-editing con Qualityity

Il *post-editing* è stato effettuato in cinque sessioni, della durata rispettiva di 02:45:57, 02:55:35, 02:31:35, 02:35:09 e 02:17:19, per un totale complessivo di 13:05:35.

Report di Post-Edit Compare

Il *post-editing* ha riguardato 190 segmenti, per un totale di circa 4.626 parole. La grande maggioranza, 172 segmenti, è stata generata automaticamente dal sistema di TA, mentre 18 segmenti risultano introdotti come *New*. Non si individuano proposte provenienti dalla memoria di traduzione, pertanto, il testo di partenza si basa interamente sui risultati della TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	19	336	7,26%
95-99%	20	600	12,97%
85-94%	51	1.577	34,09%
75-84%	33	880	19,02%
50-74%	21	366	7,91%
New (0-49%)	46	867	18,74%

Tabella 8 (All. 3): Distribuzione delle fasce di PEMp per testo n. 4 del progetto n. 2

La distribuzione delle modifiche mostra un *effort* di revisione piuttosto variegato. La fascia 100% raccoglie appena il 7,26% delle parole, indicando le porzioni accettate integralmente senza alcun intervento. Si tratta di un valore limitato, che indica come la piena affidabilità della TA sia circoscritta a un numero ristretto di segmenti. La fascia 95-99%, con il 12,97% delle parole, segnala le correzioni minime. Il nucleo più consistente delle modifiche si concentra nella fascia 85-94%, che rappresenta il 34,09% delle parole. Questo dato è estremamente significativo in quanto suggerisce che più di un terzo del testo ha richiesto interventi di lieve entità. La fascia 75-84% registra il 19,02% e indica revisioni di media entità; la fascia 50-74% con 7,91%, segnala gli interventi di *post-editing* più incisivo. Infine, la categoria *New* (0-49%) copre il 18,74% delle parole, a evidenziare un volume non trascurabile di riscritture di ampia portata fino a rielaborazioni integrali.

Complessivamente, emerge un output della TA solo parzialmente riutilizzabile: meno di un decimo del testo è stato accettato senza modifiche. La restante parte del testo è stata interessata da modifiche di intensità diversa, privilegiando in particolare le correzioni di lieve e media intensità e le rielaborazioni integrali o semi-integrali. Pertanto, la traduzione generata automaticamente non

garantisce da sola la qualità necessaria alla pubblicazione, confermando la centralità dell'intervento umano, chiamato a bilanciare piccole correzioni con riscritture più estese.

Testo n. 5 (7.950 tokens)

Nicklaus, Martina. 2020. «Wann klingt (übersetzte) Sprache fremd?» *trans-kom* 13 (2): 125–44.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	233
A/MisTra/SemUn	94
F/Sty/Disfl	78
F/Lex/LexCho	52
A/Termi/SemUn	49
A/DNT	44
A/Omis	25
A/Untra	18
A/Mech/Punct	18
A/MisTra/WSD	17
G/Orth/WhiSpa	16
F/SLP/Nomin	11
A/SouEr	9
A/MisOrd/WO	9
F/Sty/Reg	8
G/WF/Agree	6
A/Add	6
G/WF/VF	6
A/MisExpl	6
A/ForLa	6
G/Orth/Capit	5
G/Orth/Punct	4
A/MisTra/Part	4
A/MisTra/POS	4
A/MisTra/WroPr	4
F/Coh/CoRef	4
F/Lex/Fue	3
A/MWE/Komp	3
F/Sty/Idiom	2
F/De/CharFor	2
F/SLP/ForImp	1
F/Coh/Cit	1
A/Mech/CharFor	1
A/Termi/DNT	1
A/MisOrd/MW	1

Tabella 9 (All. 3): Analisi quantitativa degli errori per il testo n. 5 del progetto n. 2

Si riscontrano 395 errori di *Fluency*, 319 di *Accuracy* e 37 di *Grammar*, per un totale di 751. La maggioranza degli errori appartiene alla macrocategoria *Fluency*, principalmente a causa dell'elevata incidenza della categoria F/Coh/InPu (233 occorrenze) che segnala i casi di uso incoerente delle virgolette e dell'apostrofo. Seguono, in ordine decrescente, le seguenti categorie: A/MisTra/SemUn per gli errori di mancata corrispondenza semantica tra il testo di arrivo e il testo di partenza (94 occorrenze); F/Sty/Disfl per le problematiche di scorrevolezza della lingua di arrivo (79 occorrenze) e F/Lex/LexCho per la scelta lessicale errata in riferimento al genere testuale e all'ambito trattati (52 occorrenze). Rilevanti sono anche le occorrenze per la categoria A/DNT (44), riferibili alla presenza di numerose citazioni in lingue diverse dalla lingua di partenza e di arrivo e alla necessità di non tradurle per rispettare la scelta dell'autore di riportarli nella versione originale. Per quanto riguarda gli errori stilistici legati al linguaggio scientifico, si individuano 11 errori di nominalizzazione non realizzata (F/SLP/Nomin), 8 errori di registro inadeguato (F/Sty/Reg), 3 errori di traduzione di *Fügungen* (F/Lex/Fue) e un errore di mancato uso della forma impersonale (F/SLP/ForImp). Sul piano terminologico, vengono annotati 49 errori di traduzione di termini (A/Termi/SemUn) e un solo caso di traduzione non necessaria di termini (A/Termi/DNT). La categoria F/Coh/Cit registra una sola occorrenza, evidenziando un solo caso di strutturazione incoerente dei riferimenti bibliografici.

Durata delle sessioni di post-editing con Qualityity

Il *post-editing* è stato completato nell'arco di sei sessioni, della durata rispettiva di 02:37:21, 02:36:20, 00:50:43, 00:36:11, 04:37:31 e 02:42:11, per un totale complessivo di 14:00:17.

Report di Post-Edit Compare

Il report di Post-Edit Compare elaborato per questo testo comprende 409 segmenti (per un totale di 5.898 parole). La quasi totalità, 408 segmenti, è stata generata dal sistema di TA, mentre un solo segmento risulta introdotto come traduzione 'nuova'. Il report non evidenzia segmenti provenienti da memoria di traduzione, indicando che il testo deriva esclusivamente dalla TA.

Fasce di PEMP	Segmenti	Parole	Percentuali
100%	177	1.716	29,09%
95-99%	21	544	9,22%
85-94%	44	1.018	17,26%
75-84%	30	675	11,44%
50-74%	26	572	9,70%
New (0-49%)	111	1.373	23,28%

Tabella 10 (All. 3): Distribuzione delle fasce di PEMP per testo n. 5 del progetto n. 2

La tabella contenente le fasce di PEMp evidenzia una distribuzione eterogenea tra segmenti sottoposti a *post-editing*. La fascia 100% raccoglie il 29,09% delle parole, indicando che quasi un terzo del testo è stato riutilizzato senza alcuna modifica e mettendo in luce la buona resa del sistema su porzioni significative del documento. La fascia 95-99% rappresenta il 9,22% delle parole, comprendendo correzioni di minore portata per una porzione limitata di testo. La fascia 85-94% copre il 17,26% delle parole, evidenziando interventi di media entità. Revisioni più consistenti sono indicate dalla fascia 75-84% con 11,44% e dalla fascia 50-74% con il 9,70%, che presentano pertanto dei risultati limitati. Infine, la categoria *New* (0-49%) copre il 23,28% delle parole. Questo valore risulta significativo in quanto suggerisce che quasi un quarto del totale del testo è stato in gran parte o completamente rielaborato.

Globalmente, l'output della TA si presenta come una risorsa valida, pur mostrando discontinuità nella resa dei segmenti: una quota significativa del testo è stata accettata integralmente, tuttavia, gran parte ha richiesto interventi di varia entità, privilegiando correzioni di lieve portata e riscritture totali o quasi. La presenza in quantità significativa di segmenti 100% e *New* (0-49%) evidenzia come il lavoro del *post-editor* sia fondamentale per assicurare qualità e coerenza complessiva del documento. Dal punto di vista dell'*effort*, l'attività di *post-editing* si estende da correzioni superficiali fino a riformulazioni più estese e consistenti.

Testo n. 6 (8.513 tokens)

Schindler, Kirsten. 2021. «Literarisches Übersetzen – Eine besondere Form des Schreibens». *trans-kom* 14 (1): 43–62.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	137
F/Sty/Disfl	108
A/MisTra/SemUn	67
A/DNT	38
F/Lex/LexCho	38
A/MisTra/Gend	28
A/MisTra/Part	20
A/MisTra/WSD	20
G/Orth/Capit	20
A/Omis	12
F/Sty/Reg	12
G/Orth/Punct	12
A/Oth	12
F/Coh/Cit	10

A/SouEr	9
A/Untra	9
A/MisTra/POS	6
F/SLP/Nomin	6
A/MisOrd/WO	5
F/Coh/CoRef	5
A/Mech/Punct	4
A/ForLa	3
F/Lex/Fue	3
G/WF/Agree	3
G/Orth/Spel	3
A/MisExpl	2
A/Termi/SemUn	2
A/MisTra/WroPr	2
A/MisOrd/MW	2
A/MWE/Ell	2
G/Orth/WhiSpa	2
A/Add	1

Tabella 11 (All. 3): Analisi quantitativa degli errori per il testo n. 6 del progetto n. 2

Nell'articolo scientifico preso in esame si riscontrano 319 errori di *Fluency*, 244 errori di *Accuracy* e 40 errori di *Grammar*. Il totale degli errori annotati ammonta a 603. La quota maggiore riguarda la *Fluency*: le categorie F/Coh/InPu e F/Sty/Disfl da sole rappresentano 247 occorrenze (rispettivamente 138 e 109). Gli errori di F/Coh/InPu riguardano l'uso incoerente delle virgolette e dell'apostrofo; la categoria F/Sty/Disfl segnala le problematiche di scorrevolezza della lingua di arrivo. Per quanto riguarda gli errori di *Accuracy*, assumono rilievo le categorie di A/MisTra/SemUn e A/DNT con rispettivamente 67 e 38 errori. La categoria A/MisTra/SemUn indica la mancata corrispondenza tra il testo di partenza e il testo di arrivo. Invece, nel caso particolare degli errori di A/DNT, il numero elevato di occorrenze è dovuto alla tematica dell'articolo che richiede la presenza di numerosi esempi, la cui traduzione non è necessaria. Degna di nota è anche la categoria A/MisTra/Gend che segnala gli errori nell'uso del linguaggio inclusivo e che in questo testo registra 28 occorrenze. Con riferimento agli errori stilistici propri del linguaggio scientifico, si registrano 38 errori di lessico inadeguato all'ambito e al genere testuali trattati (F/Lex/LexCho), 12 errori di registro inappropriato (F/Sty/Reg), 6 errori di nominalizzazione non realizzata (F/SLP/Nomin) e 3 errori di traduzioni di *Fügungen* (F/Lex/Fue). Per la componente terminologica, gli errori risultano limitati a causa della scarsa presenza di terminologia specialistica, con soltanto 2 errori di traduzione errata (A/Termi/SemUn). Rilevante per il genere testuale dell'articolo scientifico è anche la categoria F/Coh/Cit che indica 10 occorrenze di strutturazione incoerente dei riferimenti bibliografici.

Durata delle sessioni di post-editing con Quality

Il *post-editing* è stato eseguito in sei sessioni, della durata rispettivamente di 02:19:09, 00:58:29, 01:25:06, 01:10:33, 03:33:29 e 02:05:17, per un totale complessivo di 11:32:03.

Report di Post-Edit Compare

Il report ha interessato 424 segmenti (per un totale di circa 6.758 parole). La quasi totalità, 422 segmenti, è stata generata dal sistema di TA, mentre 2 segmenti risultano introdotti come traduzione ex novo. Nessun segmento proviene dalla memoria di traduzione confermando la natura completamente automatica del testo di arrivo.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	141	1.822	26,96%
95-99%	58	1.224	18,11%
85-94%	71	1.533	22,68%
75-84%	39	628	9,29%
50-74%	42	571	8,45%
New (0-49%)	73	980	14,50%

Tabella 12 (All. 3): Distribuzione delle fasce di PEMp per testo n. 6 del progetto n. 2

L'analisi delle fasce di PEMp evidenzia la seguente distribuzione delle modifiche di *post-editing*. La fascia 100% raccoglie il 26,96% delle parole, indicando che oltre un quarto del testo è stato riutilizzato senza modifiche. Si tratta di un dato piuttosto significativo, che suggerisce una buona resa del sistema su porzioni significative del testo. La fascia 95-99%, riguardante le correzioni minime, copre il 18,11% delle parole. Gli interventi di media entità, invece, sono segnalati dalla fascia 85-94% che copre il 22,68% delle parole interessando quindi quasi un quarto del testo. La fascia 75-84% con il 9,29% e quella 50-74% con l'8,45% riguardano correzioni più considerevoli e registrano in questo testo dei risultati contenuti. Infine, la categoria *New* (0-49%) copre il 14,50% delle parole (73 segmenti, 980 parole), indicando una parte limitata, ma non trascurabile, di segmenti interamente o quasi riscritti.

Complessivamente, l'analisi evidenzia un output della TA che fornisce una base utile ma non completamente omogenea: una parte consistente, più di quarto, del testo è stata accettata senza modifiche, mentre la restante parte ha richiesto interventi di varia entità. Spiccano, tra gli interventi, quelli di minima e lieve entità. Non si escludono tuttavia interventi di media e medio-alta intensità. In sintesi, i risultati evidenziano l'importanza del lavoro del *post-editor* per garantire una qualità pubblicabile.

Testo n. 7 (8.484 tokens)

Agnetta, Marco. 2021. «Kompensation und Redundanz. Zwei Prinzipien der Text-Bild-Beziehung in leicht verständlichen Kommunikaten». *trans-kom* 14 (2): 257–77.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
A/MisTra/SemUn	140
F/Coh/InPu	134
F/Lex/LexCho	98
F/Sty/Disfl	74
A/Termi/SemUn	51
A/Mech/Punct	30
A/Omis	24
G/Orth/Punct	20
A/MisTra/WSD	19
A/DNT	17
G/Orth/Capit	16
A/Untra	15
A/MisTra/Part	14
A/ForLa	14
F/SLP/Nomin	12
A/MisOrd/WO	11
G/Orth/WhiSpa	10
F/Sty/Reg	9
A/Add	9
A/MisExpl	9
A/MisTra/POS	8
F/SLP/ForImp	7
A/Termi/DNT	7
A/MisOrd/MW	7
A/SouEr	7
A/Oth	6
G/WF/Agree	5
F/Coh/CoRef	4
F/Coh/Cit	3
F/Lex/Fue	3
A/Mech/CharFor	2
A/MisTra/Gend	2
F/Sty/Idiom	2
A/MisTra/WroPr	2
A/Mech/Capit	1
G/Orth/Spel	1
G/WF/VF	1

Tabella 13 (All. 3): Analisi quantitativa degli errori per il testo n. 7 del progetto n. 2

Nel testo vengono rilevati 346 errori di *Fluency*, 395 di *Accuracy* e 53 di *Grammar*, per un totale di 794 annotazioni. In questo caso, gli errori di *Accuracy* superano quelli di *Fluency*. Le categorie più ricorrenti occupano le prime quattro posizioni della tabella e sono distribuite come segue: A/MisTra/SemUn con 140 errori; F/Coh/InPu con 134 errori; F/Lex/LexCho con 98 errori; F/Sty/Disfl con 74 errori. Gli errori di A/MisTra/SemUn riguardano la mancata corrispondenza semantica tra il segmento di arrivo e quello di partenza; gli errori di F/Coh/InPu segnalano l'uso incoerente delle virgolette e dell'apostrofo; gli errori di F/Lex/LexCho individuano le scelte semantiche inadeguate all'ambito e al genere testuale trattati; infine, gli errori di F/Sty/Disfl riguardano le problematiche di scorrevolezza della lingua di arrivo. Per quanto riguarda gli aspetti stilistici legati al linguaggio scientifico, si constatano 12 errori di nominalizzazione non realizzata (F/SLP/Nomin), 9 errori di registro inappropriato (F/Sty/Reg), 7 errori di forma impersonale mancata (F/SLP/ForImp) e 3 errori di traduzione delle *Fügungen* (F/Lex/Fue). Sul piano terminologico, gli errori riscontrati ammontano a 51 occorrenze per la categoria A/Termi/SemUn, che riguarda i termini tradotti male, e 7 occorrenze per A/Termi/DNT, riferibili alla traduzione non necessaria di termini. La categoria F/Coh/Cit, pur con sole 3 annotazioni, risulta rilevante in quanto riguarda la gestione incoerente delle citazioni bibliografiche.

Durata delle sessioni di post-editing con Qualitivity

Il *post-editing* è stato svolto in sei sessioni della durata di 02:08:47, 00:47:46, 01:42:26, 02:03:42, 01:22:22 e 03:57:50, per un totale complessivo di 12:02:53.

Report di Post-Edit Compare

Il report generato da Post-Edit Compare per questo documento registra 358 segmenti, corrispondenti a circa 6.728 parole. Quasi l'intero testo, 357 segmenti, è stato generato dal sistema di TA DeepL, mentre soltanto un segmento risulta inserito manualmente come *New*. Il report non registra alcuna proposta della memoria di traduzione: mancano *matches*, segno che la base di partenza coincide integralmente con l'output prodotto dalla TA.

Fasce di PEMP	Segmenti	Parole	Percentuali
100%	121	1.831	27,21%
95-99%	16	430	6,39%
85-94%	32	849	12,62%
75-84%	20	597	8,87%
50-74%	25	608	9,04%
New (0-49%)	144	2.413	35,87%

Tabella 14 (All. 3): Distribuzione delle fasce di PEMP per testo n. 7 del progetto n. 2

La tabella 47 mostra una tendenza polarizzata degli interventi di *post-editing*. Da un lato, la fascia 100% raccoglie il 27,21% delle parole, suggerendo che più di un quarto del testo è stato accettato senza alcun intervento. Questo dato è da considerarsi significativo ed evidenzia una buona qualità dell'output in porzioni consistenti del testo. Dall'altro lato, la categoria *New* (0-49%) raggiunge un valore molto elevato, 35,87%, indicando che oltre un terzo del testo è stato riscritto integralmente. Questa dicotomia mette in luce la natura discontinua della qualità dell'output: traduzioni in alcuni casi molto affidabili, in altri totalmente inadeguate. Le fasce intermedie confermano questa variabilità: la 95-99% copre appena il 6,39% delle parole, segnalando pochi casi di correzioni minime. La 85-94% rappresenta il 12,62%, mentre le fasce 75-84% e 50-74% coprono rispettivamente l'8,87% e il 9,04% mettendo in luce la diffusione ridotta di interventi di lieve, media e medio-alta portata.

Nel complesso, il quadro suggerisce un output fortemente eterogeneo. Questo andamento mette in evidenza un aspetto cruciale: la traduzione generata di sistemi automatici non produce un livello di qualità uniforme, ma alterna risultati eccellenti a segmenti inutilizzabili senza un intervento massiccio. Questo ha comportato, dal punto di vista operativo, un lavoro sbilanciato: accanto a segmenti pronti all'uso, si sono dovute operare riscritture estese che hanno comportato un *effort* considerevole.

Testo n. 8 (9.255 tokens)

Longhi, Elisabetta. 2022. «Moorheilbäder für Italiener. Ein linguistischer Fall». *trans-kom* 15 (1): 98–127.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	226
A/DNT	102
F/Sty/Disfl	98
A/MisTra/SemUn	83
F/Lex/LexCho	64
A/Termi/SemUn	45
F/De/CharFor	37
A/Mech/Punct	33
A/Untra	22
G/Orth/WhiSpa	21
G/Orth/Capit	20
A/Omis	18
A/MisTra/WSD	18

F/Sty/Reg	18
A/MisExpl	16
A/SouEr	15
G/Orth/Punct	15
A/ForLa	11
F/Coh/Cit	7
A/MisTra/WroPr	6
F/Lex/Fue	5
F/Coh/CoRef	5
A/Termi/DNT	4
A/MisTra/Part	3
A/MisTra/POS	3
A/MisOrd/WO	3
A/MWE/Komp	3
A/MisTra/Gend	3
F/SLP/Nomin	3
A/Add	2
F/De/Layo	2
F/SLP/ForImp	2
G/WF/VF	2
G/Orth/Spel	2
A/MisOrd/MW	1
F/Sty/Idiom	1
G/WF/Agree	1

Tabella 15 (All. 3): Analisi quantitativa degli errori per il testo n. 8 del progetto n. 2

Nel testo analizzato si riscontrano 468 errori di *Fluency*, 391 errori di *Accuracy* e 61 errori di *Grammar*, per un totale complessivo di 920 annotazioni. La categoria *Fluency* risulta la più problematica, principalmente a causa della frequenza molto elevata di errori F/Coh/InPu (226 casi), riconducibili a un uso incoerente dei segni ortografici, in particolare le virgolette e l’apostrofo. Tra le altre categorie con un numero significativo di occorrenze si segnalano: A/DNT per i segmenti da non tradurre (102 occorrenze); F/Sty/Disfl per le problematiche di scorrevolezza della lingua di arrivo (98 occorrenze); e A/MisTra/SemUn per la mancata corrispondenza tra il testo di arrivo e il testo di partenza (83 occorrenze). L’alta incidenza degli errori della categoria A/DNT è direttamente legata al tema del testo: la traduzione di termini che non devono essere sottoposti a traduzione, motivo per cui il sistema genera numerose rese non necessarie. Per quanto riguarda gli aspetti stilistici del linguaggio scientifico, si registrano 64 errori di lessico non idoneo al genere testuale e all’ambito trattati (F/Lex/LexCho), 18 errori di registro inappropriato (F/Sty/Reg), 5 errori di traduzione di *Fügungen* (F/Lex/Fue), 3 errori di nominalizzazione non realizzata (F/SLP/Nomin) e 2 errori di mancato uso della forma impersonale (F/SLP/ForImp). Sul versante terminologico, si annotano 45 errori di A/Termi/SemUn (traduzione errata) e 4 errori di A/Termi/DNT (traduzione

non necessaria). Altre categorie degne di nota sono: F/De/CharFor per gli errori di formattazione sulla base delle convenzioni linguistiche del testo di arrivo (37 occorrenze) e A/Mech/Punct per le problematiche di trasferimento della punteggiatura (33 occorrenze). Infine, la categoria F/Coh/Cit registra 4 occorrenze, dato rilevante per il genere testuale trattato poiché segnala i casi di strutturazione incoerente dei riferimenti bibliografici.

Durata delle sessioni di post-editing con Qualityity

Il processo di *post-editing* è stato svolto in cinque sessioni della durata di 03:06:01, 03:15:26, 01:06:54, 03:36:27, 00:16:37 e 02:30:12, per un monte ore complessivo di 13:51:37.

Report di Post-Edit Compare

Il presente report riguarda 502 segmenti, corrispondenti a circa 7.121 parole. DeepL ha tradotto quasi la totalità del testo (501 segmenti); solo un segmento è stato introdotto ex novo. Non risultano matches dovuti all'impiego della memoria di traduzione nel progetto di traduzione, circostanza che indica come l'output iniziale sia interamente il risultato della TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	248	1.718	24,13%
95-99%	44	1.051	14,76%
85-94%	66	1.890	26,54%
75-84%	24	597	8,38%
50-74%	23	513	7,20%
New (0-49%)	97	1.352	18,99%

Tabella 16 (All. 3): Distribuzione delle fasce di PEMp per testo n. 8 del progetto n. 2

La distribuzione delle modifiche mostrata nella tabella 49 mette in luce un profilo articolato, con valori che oscillano tra segmenti accettati integralmente e casi che hanno richiesto riscritture complete. La fascia 100% rappresenta il 24,13% delle parole, indicando che quasi un quarto del testo è stato riutilizzato senza alcun intervento. La fascia 95-99% copre il 14,76%, indicando correzioni minime per una porzione non trascurabile di testo. Diversamente, la fascia 85-94% raccoglie il 26,54%, rappresentando il dato più consistente: questo significa che oltre un quarto del testo è stato interessato da modifiche di lieve intensità. Le fasce inferiori testimoniano la presenza di revisioni più consistenti ma in misura molto limitata: 8,38% delle parole nella fascia 75-84% e 7,20% per la fascia 50-74%. La categoria *New* (0-49%) assume un peso rilevante, pari al 18,99%, e segnala la necessità di riscritture radicali su quasi un quinto del testo.

Nel complesso, questo profilo evidenzia un output non uniforme: se da un lato circa un quarto del materiale è stato accettato senza correzioni, dall'altro una parte consistente (quasi un

quinto) ha richiesto un intervento completo. La restante parte ha richiesto correzioni di varia portata, con una spiccata predominanza degli interventi di lieve entità. I dati presentati mettono in luce la natura discontinua della TA, capace di produrre buoni risultati ma non di garantire coerenza e affidabilità costante.

Testo n. 9 (11.097 tokens)

Roncoroni, Tiziana. 2022. «Artikelgebrauch in Text und Übersetzung. Beobachtungen zum Sprachenpaar Deutsch-Italienisch». *trans-kom* 15 (1): 242–69.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	209
F/Sty/Disfl	95
A/MisTra/SemUn	92
A/DNT	75
F/Lex/LexCho	71
A/Untra	63
G/Orth/Capit	43
A/SouEr	36
A/Termi/SemUn	34
A/MisTra/WSD	18
F/De/Layo	17
G/Orth/Spel	16
G/Orth/Punct	15
A/Mech/Punct	15
A/ForLa	14
A/Omis	12
A/MisTra/WroPr	12
G/Orth/WhiSpa	11
F/SLP/Nomin	9
F/Sty/Reg	9
A/MisTra/Part	8
G/WF/Agree	7
A/MisTra/Gend	7
F/Lex/Fue	7
A/MisOrd/WO	5
F/Coh/CoRef	5
F/SLP/ForImp	3
A/Add	3
F/De/CharFor	3
A/MisExpl	2
A/MisTra/POS	2
A/Mech/CharFor	2

A/Termi/DNT	1

Tabella 17 (All. 3): Analisi quantitativa degli errori per il testo n. 9 del progetto n. 2

Complessivamente si individuano 921 errori, di cui 428 di *Fluency*, 401 di *Accuracy* e 92 di *Grammar*. La quantità di errori di *Fluency* risulta leggermente superiore a quella di *Accuracy*. La categoria più rappresentata è F/Coh/InPu, con 209 occorrenze, riguardante l'uso incoerente delle virgolette e dell'apostrofo. Seguono per numero di occorrenze le seguenti categorie: F/Sty/Disfl per le problematiche di scorrevolezza della lingua di arrivo (95 occorrenze); A/MisTra/SemUn per la mancata corrispondenza semantica tra il testo di arrivo e quello di partenza (92 occorrenze); A/DNT per i segmenti da non tradurre (75 occorrenze) e F/Lex/LexCho per la scelta lessicale errata in riferimento all'ambito e al genere testuale trattati (71 occorrenze). L'elevato numero di errori di A/DNT è legato alla tematica del testo, che include numerosi esempi oggetto di studio, la cui traduzione nella lingua di arrivo non è considerata necessaria. Altre categorie di errori rilevanti sono: A/Untra per i segmenti non tradotti (63 occorrenze); G/Orth/Capit per l'uso mancato della lettera maiuscola (43 occorrenze) e A/SouEr per gli errori presenti nel testo di partenza. Gli errori della categoria A/SouEr derivano da una collocazione errata del numero della nota a piè di pagina, come già evidenziato in casi precedenti. Per quanto riguarda gli errori stilistici del linguaggio scientifico, si constatano 9 errori di registro inadeguato (F/Sty/Reg), 9 errori di nominalizzazione non realizzata (F/SLP/Nomin), 7 errori di traduzione di *Fügungen* (F/Lex/Fue) e 3 errori di mancato uso della forma impersonale (F/SLP/ForImp). In riferimento alla componente terminologica, si riscontrano 34 errori di A/Termi/SemUn per i termini tradotti in modo sbagliato e solo un errore di A/Termi/DNT per la traduzione non necessaria di un termine.

Durata delle sessioni di post-editing con Qualityity

Il *post-editing* è stato portato a termine in sei sessioni della durata di 01:11:42, 03:38:16, 00:29:36, 00:29:55, 02:54:16 e 02:59:38, per un totale complessivo di 11:43:23.

Report di Post-Edit Compare

Il report di Post-Edit Compare relativo a questo documento comprende 656 segmenti, pari a circa 8.377 parole. Dei segmenti totali, 633 derivano dall'output del sistema di TA, mentre 23 sono introdotti manualmente durante il *post-editing*. Non risultano segmenti derivati dalla memoria di traduzione, a conferma come l'intera base del lavoro di *post-editing* derivi dalla TA.

Fasce di PEMP	Segmenti	Parole	Percentuali
100%	347	2.553	30,48%

95-99%	36	545	10,27%
85-94%	82	2.019	24,10%
75-84%	36	551	6,58%
50-74%	36	542	6,47%
New (0-49%)	119	1.852	22,11%

Tabella 18 (All. 3): Distribuzione delle fasce di PEMp per testo n. 9 del progetto n. 2

L'analisi delle modifiche evidenzia una tendenza polarizzata tra segmenti accettati integralmente e porzioni riscritte ex novo. La fascia 100% raggiunge infatti il 30,48% delle parole, un valore particolarmente elevato rispetto ad altri documenti analizzati: questo significa che quasi un terzo del testo è stato mantenuto invariato. All'estremo opposto, la categoria *New* (0-49%) raccoglie il 22,11%, a mostrare che oltre un quinto del testo è stato riscritto in larga misura o integralmente. Tra queste due estremità si collocano le fasce intermedie: la fascia 95-99% con interventi lievi copre il 10,27% delle parole; la 85-94% costituisce il 24,10%, indicando interventi di media entità per un quarto del totale del testo; mentre le fasce 75-84% e 50-74% rispettivamente con il 6,58% e il 6,47% segnalano revisioni più consistenti per una porzione molto limitata di testo.

Emerge complessivamente un output della TA fortemente disomogeneo: da un lato la macchina è stata in grado di produrre segmenti di qualità pubblicabile, utilizzabili senza modifiche, dall'altro lato ha generato un numero rilevante di segmenti che è stato necessario riscrivere in modo consistente o integrale. Questo sbilanciamento rivela un andamento irregolare della qualità della TA, alternando risultati eccellenti a traduzioni non funzionali. Dal punto di vista del *post-editing*, si eseguono interventi ridotti con un *effort* basso fino a riscritture integrali con un *effort* notevole.

Riflessione conclusiva

Complessivamente, le tabelle quantitative dei singoli testi mostrano che la macrocategoria di errori *Fluency* registra il numero più elevato di occorrenze, a eccezione del testo n. 7, in cui prevale invece *Accuracy*. Nella maggior parte dei testi segue, in ordine decrescente di occorrenze, la macrocategoria *Accuracy* e, infine, *Grammar*, caratterizzata dal numero più contenuto di errori. Questi dati suggeriscono che le criticità legate alla scorrevolezza della lingua di arrivo rappresentano l'ostacolo quantitativo principale per la qualità finale del testo scientifico. Per un'analisi più approfondita e qualitativa delle tipologie di errore si rimanda al capitolo 6, paragrafo 6.2.

L'analisi complessiva dei dati evidenzia, anche in questo caso, una tendenza generale di correlazione positiva tra la lunghezza dei testi, espressa in *tokens*, e la durata complessiva del *post-editing*. Generalmente, i testi più estesi richiedono tempi di correzione maggiori, come mostra il testo n. 1. Tuttavia, la relazione tra estensione del testo e tempo di *post-editing* non risulta sempre

proporzionale: il testo n. 4, ad esempio, pur avendo un numero di *tokens* contenuto (5.701), presenta un tempo di correzione superiore rispetto a testi più estesi come il testo n. 3 o il testo n. 6. Ciò suggerisce che la durata del *post-editing* non dipende esclusivamente dall'estensione del testo, ma anche da fattori qualitativi.

Un'osservazione analoga emerge dal confronto tra il numero totale di errori annotati e la durata complessiva del *post-editing*. In linea generale, testi caratterizzati da un maggior numero di errori richiedono tempi di *post-editing* più consistenti; tuttavia, anche in questo caso la tendenza presenta delle eccezioni: ad esempio, il testo n. 9 mostra un numero di errori simile a quello del testo n. 8 (rispettivamente 921 e 920), ma una durata sensibilmente inferiore (11:43:23 rispetto a 13:51:37). Questo dato lascia ipotizzare che gli errori del testo n. 9 fossero più semplici da individuare e correggere, riducendo così l'*effort* cognitivo del *post-editor*.

Il rapporto tra il numero di *tokens* e la durata del *post-editing* consente di stimare una produttività media compresa tra circa 550 e 700 *tokens* all'ora, con variazioni significative tra i testi.

Per quanto riguarda i risultati dei report di Post-Edit Compare, non emergono tendenze univoche nella distribuzione delle fasce di *post-editing*, che si presenta nel complesso piuttosto eterogenea. La fascia 85-94%, riguardante le correzioni lievi, presenta la quota più consistente in quattro testi, con valori compresi tra il 26% e il 36%. In altri tre testi, prevale la fascia 100%, con percentuali tra il 26% e il 31%, che indica i segmenti che non hanno subito modifiche. Solo in un caso la fascia più alta è 95-99% (circa 28%, correzioni minime); mentre in un altro testo si registra la predominanza della fascia *New* (circa 36%, correzioni totali o quasi). In generale, per quest'ultima categoria si individuano tre tendenze: una prima, con valori compresi tra il 13-15% (quattro testi), e una seconda con valori intorno al 18-24% (quattro testi) e un caso isolato con circa 36%. Nella fascia 100% emergono due gruppi distinti: quattro testi registrano valori tra il 9-17% e cinque testi tra il 24-31%. Le fasce 95-99% (correzioni minime), 75-84% (correzioni medie) e 50-74% (correzioni medio-alte) presentano valori estremamente variabili: per la fascia 50-74% i valori vanno dal 7-10%; per la fascia 75-84% dal 6-19%; infine per la fascia 95-99% dal 6-28%.

Questa frammentarietà dei dati rende difficile individuare un pattern coerente e suggerisce che la qualità dell'output della TA varia in modo imprevedibile tra i singoli testi. Tali dati indicano, quindi, la scarsa affidabilità del testo prodotto dalla macchina per porzioni significative di segmenti. Inoltre, l'alternanza irregolare e imprevedibile di segmenti accettabili e segmenti da modificare comporta un *effort* cognitivo elevato per il *post-editor*, che deve adattare continuamente la propria strategia di correzione alla variabilità qualitativa del testo.

Allegato 4 – Progetto n. 3 *Interculture journal*: analisi dei singoli testi

Si illustrano i risultati dell'attività di *post-editing* svolta sui testi della rivista “Interculture journal”, scelti per questo terzo progetto di traduzione. Di seguito viene riportata una tabella contenente le occorrenze per errori annotati seguita da un breve commento sulle categorie più significative; le indicazioni temporali delle sessioni di *post-editing* e il report di Post-Edit Compare presentato in una tabella che mostra le percentuali di intervento e un breve commento.

Testo n. 1 (9.135 tokens)

Nazarkiewicz, Kirsten. 2013. «Interkulturalität als immanenter Faktor in Coaching und Training – konzeptionelle Überlegungen». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 12 (20): 47–67. <https://doi.org/10.24403/jp.1285138>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	161
A/MisTra/SemUn	128
F/Lex/LexCho	94
F/Sty/Disfl	84
A/Omis	38
A/MWE/Komp	36
F/De/Layo	35
A/MisTra/Part	30
F/SLP/ForImp	21
F/SLP/Nomin	21
A/MisTra/WSD	19
G/Orth/Punct	18
A/SouEr	16
G/Orth/Capit	16
A/MisOrd/MW	15
A/DNT	15
F/Sty/Reg	13
A/MisTra/WroPr	12
A/MisOrd/WO	12
A/Add	12
A/MisTra/POS	11
F/Lex/Fue	11
A/Untra	10
F/Coh/CoRef	8
A/Mech/Punct	6
F/Sty/Idiom	6
G/Orth/WhiSpa	6

G/Orth/Spel	5
A/Mech/CharFor	4
A/ForLa	3
F/Coh/Cit	2
F/De/CharFor	2
A/MisTra/Gend	2
A/Termi/DNT	1
A/Oth	1
G/WF/VF	1

Tabella 1 (All. 4): Analisi quantitativa degli errori per il testo n. 1 del progetto n. 3

Nel testo preso in esame si riscontrano 458 errori di *Fluency*, 371 errori di *Accuracy* e 46 errori di *Grammar*. Il totale ammonta a 875 errori annotati. Gli errori di *Fluency* costituiscono la maggioranza principalmente a causa dell'elevato numero di errori di F/Coh/InPu (161 occorrenze) che segnalano l'uso incoerente delle virgolette e dell'apostrofo. Assumono un ruolo rilevante anche gli errori di corrispondenza semantica tra il testo di arrivo e quello di partenza (A/MisTra/SemUn, 128 occorrenze), gli errori di scelta lessicale inadeguata all'ambito e al genere testuale trattato (F/Lex/LexCh, 94 occorrenze) e gli errori di scorrevolezza della lingua di arrivo (F/Sty/Disfl, 84 occorrenze). Altre categorie con un numero significativo di occorrenze comprendono: A/Omis per gli errori di omissione (38 occorrenze), A/MWE/Komp per gli errori di traduzione di aggettivi composti (36 occorrenze) e F/De/Lay per gli errori di strutturazione del testo (35 occorrenze). Per quanto attiene gli errori stilistici del linguaggio scientifico, si evidenziano 21 errori di nominalizzazione non realizzata (F/SLP/Nomin), 21 uso mancato della forma impersonale (F/SLP/ForImp), 13 errori di registro inadeguato (F/Sty/Reg) e 11 errori di traduzione di *Fügungen* (F/Lex/Fue). La componente terminologica presenta un solo errore di traduzione non necessaria (A/Termi/DNT), indice della limitata presenza di termini specialistici nel testo. Infine, significativa per il genere testuale dell'articolo scientifico è la categoria F/Coh/Cit. Si tratta di errori riguardanti la strutturazione incoerente dei riferimenti bibliografici e in questo testo se ne registrano solo 2 casi.

Durata delle sessioni di post-editing con Qualitivity

Il *post-editing* si è articolato in otto sessioni, della durata rispettivamente di 01:18:16, 02:16:18, 03:35:02, 00:31:34, 03:11:15, 00:43:25, 02:10:55 e 00:55:46, per un totale complessivo di 14:42:31.

Report di Post-Edit Compare

L'analisi ha riguardato 472 segmenti per un totale di circa 7.834 parole. La maggior parte dei segmenti, 469, è il risultato della TA, mentre 3 segmenti sono aggiunti manualmente come *New*. Non si rilevano proposte della memoria di traduzione, pertanto il testo di arrivo si basa interamente

sulle proposte formulate automaticamente. Di seguito la tabella 53 riporta la distribuzione delle percentuali di PEM.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	134	1.502	19,17%
95-99%	54	1.034	13,20%
85-94%	90	2.109	26,92%
75-84%	46	979	12,50%
50-74%	33	570	7,28%
New (0-49%)	115	1.640	20,93%

Tabella 2 (All. 4): Distribuzione delle fasce di PEMp per testo n. 1 del progetto n. 3

L'esame delle fasce di PEMp mostra una distribuzione eterogenea degli interventi di *post-editing*. La fascia 100% raccoglie il 19,17% delle parole e questo suggerisce che circa un quinto del testo è stato accettato senza alcuna modifica. La fascia 95-99% rappresenta il 13,20% delle parole, corrispondente a correzioni minime. La fascia 85-94% presenta la quota più elevata, coprendo il 26,92% delle parole e segnalando che oltre un quarto del testo è stato coinvolto da interventi di media entità. Gli interventi di *post-editing* più consistenti sono rappresentati dalle fasce 75-84%, che registra il 12,50% e 50-74%, che copre il 7,28%. Questi dati indicano una diffusione limitata di questi tipi di interventi. Infine, la categoria *New* (0-49%) copre il 20,93% delle parole e indica interventi che hanno comportato rielaborazioni di larga portata o integrali. Si tratta di una percentuale considerevole che equivale a un quinto dell'intero testo.

Nel complesso, il profilo che emerge da questa analisi descrive un output della TA affidabile come punto di partenza ma non omogeneo: alla parte significativa del testo mantenuta invariata (quasi il 20%) si contrappone una parte altrettanto significativa di segmenti che sono stati interessati da rielaborazioni totali o quasi totali (quasi il 21%). I restanti segmenti hanno richiesto interventi di varia entità, con una predominanza degli interventi di lieve entità. Queste osservazioni mettono in luce l'importanza del lavoro del *post-editor* per garantire la qualità desiderata e per trasformare l'output automatico in un testo pronto per la pubblicazione.

Testo n. 2 (9.173 tokens)

Kriegel-Schmidt, Katharina, e Klaus Schmidt. 2014. «Kultur als Blickwinkel in interkulturell orientierten Lernsettings. Didaktische und theoretische Reflexionen am Beispiel einer Ausbildung in Interkultureller Mediation». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 13 (22): 27–44. <https://doi.org/10.24403/jp.1282887>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	144
A/MisTra/SemUn	120
F/Sty/Disfl	107
G/Orth/Capit	81
F/Lex/LexCho	60
F/Sty/Reg	26
A/Omis	24
F/SLP/ForImp	24
A/MisTra/Part	21
A/MisTra/WSD	20
F/SLP/Nomin	19
F/Coh/CoRef	19
G/Orth/Punct	18
A/Untra	17
G/Orth/WhiSpa	14
F/Lex/Fue	13
A/Add	11
G/Orth/Spel	10
A/MisTra/POS	8
F/De/Layo	7
G/WF/VF	7
A/MisOrd/WO	7
A/MWE/Komp	6
A/DNT	5
A/SouEr	5
A/Termi/SemUn	4
F/Sty/Idiom	4
A/Mech/CharFor	3
A/MisTra/WroPr	2
A/MisExpl	1
A/MisOrd/MW	1
A/MisTra/Gend	1
A/Oth	1
G/WF/Agree	1

Tabella 3 (All. 4): Analisi quantitativa degli errori per il testo n. 2 del progetto n. 3

Nel corso dell'analisi del testo vengono individuati 423 errori di *Fluency*, 257 errori di *Accuracy* e 131 errori di *Grammar*, per un totale complessivo di 811 annotazioni. La maggioranza degli errori riguarda la *Fluency*, principalmente a causa della categoria F/Coh/InPu, che registra 144 occorrenze e segnala i casi di uso incoerente delle virgolette e dell'apostrofo. Tra le categorie con maggiore incidenza si riscontrano anche A/MisTra/SemUn per gli errori di mancata corrispondenza semantica tra il testo di arrivo e quello di partenza (120 occorrenza); F/Sty/Disfl per le problematiche di scorrevolezza della lingua di arrivo (107 occorrenze); G/Orth/Capit per l'uso

grammaticalmente errato della lettera maiuscola (80 occorrenze) e di F/Lex/LexCh per la scelta lessicale errata in riferimento all'ambito e al genere testuale trattati (60 occorrenze). Quest'ultima categoria è particolarmente rilevante per la valutazione della resa del linguaggio scientifico. Tra gli altri errori riguardanti lo stile scientifico, si rilevano inoltre 26 errori di registro inadeguato (F/Sty/Reg), 24 di mancato uso della forma impersonale (F/SLP/ForImp), 19 errori di nominalizzazione non realizzata (F/SLP/Nomin) e 13 errori di traduzione di *Fügungen* (F/Lex/Fue). In riferimento alla traduzione della terminologia, vengono annotati solo 4 errori di A/Termi/SemUn per la traduzione errata di termini.

Durata delle sessioni di post-editing con Qualityity

La fase di *post-editing* è stata articolata in sei sessioni, della durata rispettivamente di 02:23:45, 04:19:25, 00:46:24, 03:55:33, 02:16:57 e 01:35:40. Il totale delle sessioni di *post-editing* ammonta a 15:17:44.

Report di Post-Edit Compare

Il report di Post-Edit Compare relativo a questo testo prende in esame 439 segmenti, per un totale di circa 7.872 parole. L'intero documento, fatta eccezione per un singolo segmento introdotto manualmente come *New*, è stato generato dal sistema di TA. La memoria di traduzione non è intervenuta mediante la presentazione di *matches*, indicando che l'analisi del testo di arrivo si basa esclusivamente sull'output della TA.

Fasce di PEMP	Segmenti	Parole	Percentuali
100%	133	2.325	29,54%
95-99%	45	1.008	12,80%
85-94%	47	1.183	15,03%
75-84%	22	424	5,39%
50-74%	19	405	5,14%
New (0-49%)	173	2.527	32,10%

Tabella 4 (All. 4): Distribuzione delle fasce di PEMP per testo n. 2 del progetto n. 3

La tabella 55 evidenzia un quadro polarizzato. Quasi un terzo del testo, o più precisamente il 29,54%, è stato accettato senza modifiche (fascia 100%). Questo dato suggerisce la capacità del sistema di TA di elaborare traduzioni di qualità e pubblicabile per una parte significativa del testo. I segmenti interessati da correzioni lievi (fascia 95-99%) e quelli coinvolti da interventi di media entità (fascia 85-94%) costituiscono rispettivamente il 12,80% e il 15,03%. Si tratta di risultati limitati ma sicuramente non trascurabili. Gli interventi di *post-editing* più consistenti sono rappresentati dalle fasce 75-84% e 50-74% che registrano rispettivamente il 5,39% e il 5,14%,

suggerendo una distribuzione limitata di questo tipo di interventi. Infine, la categoria *New* (0-49%) rappresenta il 32,10% delle parole. Questo è il dato più consistente, che indica che un terzo delle parole è stato sottoposto a rielaborazioni di ampia portata o integrali.

Nel complesso, l'output della TA può essere considerato una base di partenza solida, ma con una qualità non uniforme: alcune porzioni del testo vengono accettate senza modifiche, mentre molte altre hanno richiesto degli interventi. Le percentuali elevate per categorie 100% e *New* (0-49%) sottolineano quanto sia determinante il lavoro di *post-editing* per assicurare qualità, uniformità e leggibilità complessiva del documento. Dal punto di vista dell'*effort*, l'attività di revisione ha incluso modifiche di varia entità, con una chiara predominanza delle rielaborazioni integrali o pressoché integrali.

Testo n. 3 (7.218 tokens)

Henze, Jürgen. 2016. «Vom Verschwinden des (Inter) Kulturellen und Überleben der (Inter)Kulturalität». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 15 (26): 59–74. <https://doi.org/10.24403/jp.1278515>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	181
A/MisTra/SemUn	76
F/Sty/Disfl	63
F/Lex/LexCho	44
A/DNT	44
A/MisTra/WSD	18
A/Omis	17
F/De/Layo	17
A/Untra	16
F/SLP/Nomin	12
G/Orth/Capit	12
G/Orth/Punct	11
A/MisTra/Part	9
G/WF/Agree	9
F/Sty/Reg	7
A/Termi/SemUn	7
G/Orth/WhiSpa	7
F/SLP/ForImp	6
F/Lex/Fue	6
G/Orth/Spel	5
A/Add	5
A/MWE/Komp	4

A/MisOrd/WO	4
A/MisTra/POS	4
A/SouEr	2
A/Mech/Punct	2
A/MisTra/Gend	2
A/MisExpl	2
F/Coh/CoRef	1
F/Sty/Idiom	1
A/MWE/Ell	1
A/ForLa	1
A/Oth	1

Tabella 5 (All. 4): Analisi quantitativa degli errori per il testo n. 3 del progetto n. 3

Nel corso dell'analisi si riscontrano 338 errori di *Fluency*, 215 errori di *Accuracy* e 44 errori di *Grammar*, per un totale complessivo di 597 annotazioni. La macrocategoria *Fluency* risulta la più ingente, soprattutto per l'elevata concentrazione di errori di tipo F/Coh/InPu, che ammontano in totale a 181 occorrenze e che riguardano l'uso incoerente delle virgolette e dell'apostrofo. Seguono, per incidenza, le seguenti categorie: A/MisTra/SemUn per gli errori di corrispondenza semantica tra il testo di partenza e quello di arrivo (76 occorrenze); F/Sty/Disfl per le anomalie di scorrevolezza della lingua di arrivo (63 occorrenze); F/Lex/LexCho per la scelta lessicale errata in riferimento all'ambito e al genere testuale trattati (44 errori) e A/DNT per i segmenti non da sottoporre a traduzione (44 errori). Gli errori di quest'ultima categoria sono dovuti alla considerevole presenza di citazioni in lingua inglese e alla necessità di rispettare la volontà dell'autore riportandole in lingua inglese anche nel testo di arrivo. Per quanto riguarda gli aspetti stilistici legati al linguaggio scientifico, emergono 12 errori di nominalizzazione non realizzata (F/SLP/Nomin), 7 errori di registro inadeguato (F/Sty/Reg), 6 errori di traduzione di *Fügungen* (F/Lex/Fue) e 6 errori di mancato uso della forma impersonale (F/SLP/ForImp). Sul versante terminologico, le annotazioni risultano limitate: solo 7 errori per la categoria A/Termi/SemUn, segnalante la traduzione errata di errori.

Durata delle sessioni di post-editing con Qualitivity

Il *post-editing* si è svolto in cinque sessioni della durata di 04:16:04, 02:40:38, 02:35:46, 01:53:57, 00:20:12, per un totale complessivo di 11:46:37.

Report di Post-Edit Compare

Il report relativo a questo testo comprende 247 segmenti, pari a circa 7.872 parole. Tutti i segmenti sono prodotti dal sistema di TA, senza alcun contributo della memoria di traduzione.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	82	1.304	21,64%
95-99%	22	689	11,44%
85-94%	57	1.778	29,51%
75-84%	16	462	7,67%
50-74%	16	431	7,15%
New (0-49%)	54	1.361	22,59%

Tabella 6 (All. 4): Distribuzione delle fasce di PEMp per testo n. 3 del progetto n. 3

La distribuzione dei risultati pone in evidenza una tendenza traduttiva disomogenea. Poco più di un quinto del testo, precisamente il 21,64%, non ha subito alcuna modifica, suggerendo una resa automatica soddisfacente in alcune sezioni del testo. Una quota più ridotta, pari al 11,44%, rientra nella fascia 95-99% che rappresenta le correzioni minime. La fascia 85-94% si distingue come la più rappresentata con il 29,51% e indica gli interventi di *post-editing* di lieve portata che hanno interessato quasi un terzo del testo. Le fasce di interventi medi (75-84%) e medio-alti (50-74%) coprono rispettivamente il 7,67% e il 7,15% del totale e mostrano una diffusione piuttosto limitata. Infine, la categoria *New* (0-49%) presenta un risultato significativo, il 22,59%, segnalando la necessità di riscritture complete o quasi complete per oltre un quinto del testo.

Nel complesso, il materiale di partenza fornito dalla TA ha dimostrato di essere una base utile, ma la qualità si è rivelata discontinua: accanto a segmenti perfettamente riutilizzabili, altri hanno richiesto correzioni di varia portata o interventi radicali. La quota consistente di segmenti 100% è bilanciata dalla percentuale altrettanto elevata di segmenti *New* (0-49%), evidenziando quindi il ruolo essenziale del *post-editor* nel garantire una qualità omogenea. In termini di *effort*, l'impegno richiesto si è articolato in modo diversificato, con una porzione consistente di segmenti coinvolti da interventi di lieve entità.

Testo n. 4 (6.492 tokens)

Bolten, Jürgen. 2016. «Interkulturelle Trainings neu denken». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 15 (26): 75–91. <https://doi.org/10.24403/jp.1278516>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	131
A/MisTra/SemUn	103
F/Sty/Disfl	82
F/Lex/LexCho	64
G/Orth/Capit	36
A/Omis	26

G/Orth/WhiSpa	22
A/MisTra/WSD	21
G/Orth/Punct	16
F/De/Layo	14
A/DNT	11
A/Untra	11
A/MisOrd/WO	10
A/MisTra/POS	10
A/SouEr	10
F/SLP/ForImp	8
G/WF/Agree	8
A/Add	8
F/Sty/Reg	7
A/Mech/CharFor	7
A/Mech/Punct	7
A/MisTra/Part	6
F/SLP/Nomin	6
A/MisOrd/MW	6
A/Termi/SemUn	5
G/Orth/Spel	5
A/ForLa	5
A/MWE/Komp	4
A/MWE/Ell	4
A/Termi/DNT	3
A/MisTra/WroPr	2
A/Mech/Capit	2
A/MisTra/Gend	1
F/Sty/Idiom	1
F/De/CharFor	1
A/Oth	1
F/Coh/Cit	1

Tabella 7 (All. 4): Analisi quantitativa degli errori per il testo n. 4 del progetto n. 3

Nel complesso si annotano 665 errori, distribuiti nel seguente modo: 315 di *Fluency*, 263 di *Accuracy* e 87 di *Grammar*. La *Fluency* rappresenta la categoria quantitativamente più considerevole, superando nettamente le altre due. In ordine decrescente per numero di occorrenze, le categorie di errori più rappresentate sono le seguenti: F/Coh/InPu per i casi di uso incoerente di virgolette e apostrofo (131 errori); A/MisTra/SemUn per gli errori di corrispondenza semantica tra il testo di arrivo e il testo di partenza (103 errori); F/Sty/Disfl per le problematiche di scorrevolezza della lingua di arrivo (82 errori); F/Lex/LexCho per gli errori di scelta lessicale in riferimento all'ambito e al genere testuali trattati (64 errori). Per quanto riguarda il linguaggio scientifico, si constatano 8 errori di forma impersonale non impiegata (F/SLP/ForImp), 7 errori di registro

inadeguato (F/Sty/Reg), 6 errori di nominalizzazione non realizzata (F/SLP/Nomin). Sul piano terminologico, si riscontrano 5 errori per la categoria A/Termi/SemUn, che segnala le traduzioni errate, e 3 errori di A/Termi/DNT, riguardanti i casi di traduzione non necessaria. Significativo è anche il numero di occorrenze per la categoria G/Orth/Capit, che indica gli usi grammaticalmente errati della lettera maiuscola. Si registra infine un solo errore di strutturazione incoerente dei riferimenti bibliografici (F/Coh/Cit).

Durata delle sessioni di post-editing con Qualityity

Il *post-editing* si è svolto in quattro sessioni della durata rispettiva di 01:05:43, 04:44:38, 02:16:12, 03:53:02, per un totale complessivo di 11:59:35.

Report di Post-Edit Compare

Secondo i dati del report elaborato da Post-Edit Compare, il testo oggetto di analisi si compone di 296 segmenti, equivalenti a circa 5.253 parole. Tutti i segmenti provengono dalla TA e non si rilevano proposte della memoria di traduzione.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	85	1.217	23,17%
95-99%	28	758	14,43%
85-94%	52	1.223	23,28%
75-84%	21	433	8,24%
50-74%	18	380	7,23%
New (0-49%)	92	1.242	23,64%

Tabella 8 (All. 4): Distribuzione delle fasce di PEMp per testo n. 4 del progetto n. 3

L'analisi delle fasce di PEMp, presentata nella tabella 59, mette in evidenza risultati contrastanti. Il 23,17% delle parole, cioè quasi un quarto, è stato accettato senza alcuna modifica, segnalando un buon livello qualitativo in diverse porzioni. Gli interventi di entità minima coprono il 14,43%. Più considerevole risulta, invece, la percentuale registrata dalla fascia 85-94%, il 23,28%, che indica che quasi un quarto del testo ha richiesto modifiche di lieve intensità. Non mancano inoltre segmenti che hanno richiesto modifiche più consistenti: si tratta delle fasce 75-84% (8,24%) e 50-74% (7,23%), che registrano però percentuali contenute. A ciò si aggiunge la categoria *New* (0-49%), che con il 23,64% delle parole evidenzia una parte, quasi un quarto del testo, coinvolta in rielaborazioni di larga scala o integrali, a conferma di una qualità non uniforme dell'output.

In sintesi, il testo prodotto dalla TA si è rivelato utile come base di partenza, ma ha mostrato forti discontinuità: la percentuale alta per la fascia 100% è controbilanciata da una percentuale altrettanto alta per la fascia *New* (0-49%). I segmenti restanti hanno richiesto delle modifiche di

varia entità, privilegiando in particolar modo le modifiche lievi (85-94%). Alcune sezioni hanno richiesto interventi minimi, altre invece sono state oggetto di riscritture totali. Dall'analisi emerge, dunque, quanto il contributo del *post-editor* sia determinante per ottenere un documento coerente, leggibile e pubblicabile.

Testo n. 5 (8.706 tokens)

Barmeyer, Christoph I., e Eric Davoine. 2016. «Konstruktives interkulturelles Management – Von der Aushandlung zur Synergie». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 15 (26): 97–115. <https://doi.org/10.24403/jp.1278518>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	163
F/Sty/Disfl	113
A/MisTra/SemUn	91
A/Termi/DNT	40
F/Lex/LexCho	37
G/Orth/Capit	32
A/Omis	24
A/Untra	21
G/Orth/Punct	20
A/DNT	19
G/Orth/WhiSpa	14
A/MisTra/Part	14
A/MisTra/WSD	13
A/MisTra/POS	12
A/MisOrd/WO	10
A/Mech/Punct	10
A/Termi/SemUn	9
A/Add	8
F/Lex/Fue	7
A/MisTra/WroPr	6
A/MWE/Komp	5
F/De/Layo	5
A/SouEr	3
G/Orth/Spel	3
G/WF/Agree	3
A/Mech/CharFor	3
F/Coh/CoRef	3
F/Sty/Idiom	3
F/SLP/ForImp	2
A/MisExpl	2

F/Sty/Reg	1
F/SLP/Nomin	1
F/Coh/Cit	1
A/MWE/Ell	1
A/MisOrd/MW	1
A/Oth	1

Tabella 9 (All. 4): Analisi quantitativa degli errori per il testo n. 5 del progetto n. 3

Vengono riscontrati 336 errori di *Fluency*, 293 errori di *Accuracy* e 72 errori di *Grammar*, per un totale che ammonta a 701 errori. La macrocategoria con il maggior numero di occorrenze è la *Fluency* e questo è dovuto in particolar modo all'elevato numero di errori per la categoria F/Coh/InPu (163 occorrenze), che riguarda l'uso incoerente di virgolette e apostrofo. Seguono con frequenze particolarmente elevate le categorie F/Sty/Disfl e A/MisTra/SemUn con rispettivamente 113, 91 errori. La prima segnala le problematiche di scorrevolezza nella lingua di arrivo; la seconda viene impiegata per gli errori di corrispondenza semantica tra il segmento di arrivo e quello di partenza. Significativa è anche la presenza di errori di G/Orth/Capit (32 occorrenze), che indicano gli usi grammaticalmente scorretti della lettera maiuscola. Per quanto riguarda gli aspetti stilistici legati al linguaggio scientifico, si registrano 37 errori di scelta lessicale inadeguata al genere testuale e all'ambito trattati (F/Lex/LexCho), 12 errori di traduzione di *Fügungem* (F/Lex/Fue), 3 errori di forma impersonale non impiegata (F/SLP/ForImp), un errore di nominalizzazione non realizzata (F/SLP/Nomin) e un errore di registro inappropriato (F/Sty/Reg). Nella componente terminologica spiccano i 40 errori di A/Termi/DNT, riconducibili all'uso consolidato di termini inglesi per i quali la traduzione non è necessaria, e 9 errori di A/Termi/SemUn, relativi alla traduzione errata di termini. La categoria F/Coh/Cit registra solo un caso di strutturazione incoerente dei riferimenti bibliografici.

Durata delle sessioni di post-editing con Qualitivity

Il *post-editing* si è svolto in cinque sessioni della durata rispettiva di 02:28:21, 04:06:17, 04:50:19, 01:55:48, 04:35:56. Il totale complessivo è di 17:56:41.

Report di Post-Edit Compare

Il testo esaminato si compone di 431 segmenti, per un totale di circa 7.174 parole. La grande maggioranza, 428 segmenti, è stata prodotta dal sistema di TA, mentre soltanto 3 segmenti sono introdotti come 'nuova' traduzione. Non si rilevano segmenti provenienti dalla memoria di traduzione, a sottolineare che il testo analizzato deriva unicamente dalla TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	144	1.436	20,02%
95-99%	64	1.572	21,91%
85-94%	77	1.902	26,51%
75-84%	24	406	5,66%
50-74%	25	494	6,89%
New (0-49%)	97	1.364	19,01%

Tabella 10 (All. 4): Distribuzione delle fasce di PEMp per testo n. 5 del progetto n. 3

Nella tabella 61 è possibile osservare la distribuzione delle fasce di PEMp. La fascia 100% raccoglie il 20,02% delle parole: questo dato suggerisce che una parte significativa, un quinto del testo, è stata accettata senza alcuna modifica. La fascia 95-99% si attesta al 21,91%, con interventi minimi, mentre la fascia 85-94%, che rappresenta il 26,51%, risulta la più ingente e riflette correzioni di media entità. Le fasce intermedie, riguardanti interventi più significativi ed estesi, mostrano invece dati limitati: 75-84% con il 5,66% e 50-74% con il 6,89%. Infine, la categoria *New* (0-49%) incide in maniera rilevante sul totale, con il 19,01% delle parole, segnalando che quasi un quinto del testo ha richiesto rielaborazioni di ampia portata o addirittura integrali.

Nel complesso, si può constatare che l'output della TA fornisce una base utile ma non sempre uniforme: da un lato, un quinto del testo è stato mantenuto inalterato, dall'altro una parte quasi equivalente è stata riscritta interamente, o quasi. Le restanti parti di testo sono state interessate prevalentemente da modifiche di entità minima e lieve, come dimostrano le percentuali per le fasce 95-99% e 85-94%. Si può dedurre il ruolo chiave del *post-editor* per assicurare una qualità pubblicabile, mentre le sue attività di correzione si articolano su più livelli, da correzioni minime fino a riformulazioni più estese.

Testo n. 6 (10.157 tokens)

Gruber, Valerie V. V., e Eberhard Rothfuß. 2016. «Interkulturelle Managementforschung – Reflexive Gedanken über eine unreflektierte Denkschule». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 15 (26): 117–40. <https://doi.org/10.24403/jp.1278519>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	221
F/Lex/LexCho	94
A/MisTra/SemUn	87
F/Sty/Disfl	78
G/Orth/Punct	29
F/Coh/Cit	28

A/Omis	23
A/MisTra/WSD	21
A/MisTra/Part	20
F/Sty/Reg	15
A/Mech/Punct	14
F/SLP/Nomin	12
G/Orth/WhiSpa	11
A/Untra	11
A/MisTra/Gend	10
F/Lex/Fue	10
A/DNT	9
G/Orth/Capit	9
A/MisOrd/WO	8
A/MisTra/POS	7
A/Add	7
A/SouEr	6
F/Coh/CoRef	6
G/WF/Agree	5
G/Orth/Spel	5
A/MisTra/WroPr	4
A/Termi/SemUn	3
A/MisOrd/MW	2
A/Oth	2
A/ForLa	2
F/Sty/Idiom	2
F/De/Layo	2
F/SLP/ForImp	1
G/WF/VF	1

Tabella 11 (All. 4): Analisi quantitativa degli errori per il testo n. 6 del progetto n. 3

Si rilevano complessivamente 765 errori, suddivisi in 469 errori di *Fluency*, 236 di *Accuracy* e 60 di *Grammar*. Gli errori di *Fluency* costituiscono più della metà del totale, con una presenza particolarmente significativa della categoria F/Coh/InPu, che conta 221 casi di uso incoerente delle virgolette e della punteggiatura. Tra le altre categorie più frequenti si segnalano F/Lex/LexCho per gli errori di scelta lessicale in riferimento all'ambito e al genere testuale trattati (94 errori), A/MisTra/SemUn per gli errori di corrispondenza semantica tra il testo di partenza e quello di arrivo (87 errori) e F/Sty/Disfl per le problematiche di scorrevolezza della lingua di arrivo (78 errori). Per quanto riguarda gli aspetti stilistici propri del linguaggio scientifico, oltre agli errori di F/Lex/LexhCho, si registrano 15 errori di registro inadeguato (F/Sty/Reg), 12 errori di mancata realizzazione della nominalizzazione (F/SLP/Nomin), 10 errori di traduzione di *Fügungen* (F/Lex/Fue) e uno per la forma impersonale non impiegata (F/SLP/ForImp). Gli errori relativi alla terminologia risultano contenuti, con appena 3 errori di traduzione inesatta (A/Termi/SemUn). Per

la macrocategoria *Grammar* rilevanti sono i 29 errori individuati per G/Orth/Punct. Questi fanno riferimento all'uso grammaticalmente errato della punteggiatura. Infine, si individuano 28 occorrenze per la categoria F/Coh/Cit. Questi errori, rilevanti per il linguaggio scientifico e il genere testuale dell'articolo scientifico, riguardano la strutturazione incoerente dei riferimenti bibliografici.

Durata delle sessioni di post-editing con Qualityity

Sono state eseguite sei sessioni di *post-editing* della durata rispettiva di 02:41:09, 02:31:07, 03:09:46, 03:14:32, 02:04:27, 01:33:03, per un totale complessivo di 15:14:04.

Report di Post-Edit Compare

Il report di Post-Edit Compare relativo a questo testo prende in esame 301 segmenti, per un totale di circa 8.286 parole. La quasi totalità dei segmenti, 296, è prodotta dal sistema di TA, mentre 5 segmenti sono introdotti manualmente come *New*. Non si sono riscontrate proposte provenienti dalla memoria di traduzione confermando che l'analisi si basa esclusivamente sull'output della TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	44	864	10,21%
95-99%	64	2.099	25,33
85-94%	91	2.899	34,99%
75-84%	13	444	5,36%
50-74%	21	711	8,58%
New (0-49%)	68	1.287	15,53%

Tabella 12 (All. 4): Distribuzione delle fasce di PEMp per testo n. 6 del progetto n. 3

La distribuzione delle fasce di PEMp nella Tabella 63 mostra un quadro eterogeneo. La fascia 100% raccoglie il 10,21% delle parole, a indicare che solo una porzione limitata del testo è stata accettata senza alcuna modifica. Segmenti con correzioni minime, compresi nella fascia 95-99%, rappresentano il 25,33% delle parole, cioè un quarto del testo. La fascia 85-94% costituisce la porzione più ampia del testo, con 34,99% delle parole, suggerendo che oltre un terzo del testo è stato interessato da interventi di *post-editing* di lieve entità. Le percentuali osservate per le fasce medie e medio-alte, 75-84% con il 5,36% e 50-74% con l'8,58%, indicano che le correzioni più sostanziali hanno interessato una porzione limitata del testo. Infine, la categoria *New* (0-49%) copre il 15,53% delle parole, segnalando che una parte non trascurabile ha richiesto riscritture totali o in larga misura.

In sintesi, l'output della TA ha fornito una base di partenza utile e questo si può dedurre dalla concentrazione di interventi nelle fasce di modifiche minime e lievi. La TA, quindi, può essere di supporto ma l'intervento dell'essere umano è ancora indispensabile. In questo testo, la

percentuale di segmenti 100% è troppo bassa e non riesce a controbilanciare la fascia *New* (0-49%), che richiede correzioni ampie. In termini di *effort*, questi dati si traducono in interventi di varia entità, che il *post-editor* deve abilmente modulare a seconda dei singoli casi presi in esame.

Testo n. 7 (8.901 tokens)

Ivanova, Alina, Markus Kollmannsberger, e Ewald Kiel. 2017. «Herausfordernde Situationen im interkulturellen schulischen Kontext: Wahrnehmungs- und Handlungsmuster von Lehrkräften». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 16 (27–28): 23–41. <https://doi.org/10.24403/jp.1278337>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	298
F/Lex/LexCho	84
F/Sty/Disfl	67
A/MisTra/SemUn	65
A/MisTra/Gend	60
F/SLP/Nomin	25
A/Omis	19
A/MisTra/WSD	17
F/De/Layo	14
G/Orth/Capit	13
A/MisTra/Part	12
F/Coh/Cit	12
G/Orth/Punct	11
A/Mech/Punct	10
A/DNT	8
A/Untra	8
A/MisTra/WroPr	8
A/MisExpl	8
A/MisOrd/MW	8
A/SouEr	8
A/MisTra/POS	8
G/Orth/Spel	8
F/De/CharFor	7
A/Add	6
F/Sty/Reg	6
G/WF/Agree	4
F/Coh/CoRef	4
A/Termi/SemUn	3
F/Lex/Fue	3
A/Mech/Capit	3

A/MWE/Komp	2
A/Termi/DNT	2
G/WF/VF	2
A/MisOrd/WO	1
F/Sty/Idiom	1
A/Mech/CharFor	1
A/MWE/Ell	1
A/Oth	1

Tabella 13 (All. 4): Analisi quantitativa degli errori per il testo n.7 del progetto n. 3

Nel corso dell'analisi del testo si identificano 521 errori di *Fluency*, 259 errori di *Accuracy* e 38 errori di *Grammar*, per un totale di 818 annotazioni. Gli errori di *Fluency* costituiscono la quota più rilevante, rappresentando oltre due terzi del totale, con particolare incidenza della categoria F/Coh/InPu, osservata 298 volte, che segnala un uso incoerente delle virgolette e dell'apostrofo. Altre categorie significative includono: F/Lex/LexCho per scelta lessicale inadeguata all'ambito e al genere testuale trattati (84 occorrenze); F/Sty/Disfl per le problematiche di scorrevolezza della lingua di arrivo (67 occorrenze); A/MisTra/SemUn per la mancata corrispondenza semantica tra il testo di arrivo e quello di partenza (65 occorrenze) e A/MisTra/Gend per il mancato uso del linguaggio inclusivo (60 occorrenze). Per quanto riguarda gli aspetti stilistici legati al linguaggio scientifico, si riscontrano 25 errori di nominalizzazione non realizzata (F/SLP/Nomin), 6 errori di registro inappropriato (F/Sty/Reg) e 3 errori di traduzioni di *Fügungen* (F/Lex/Fue). Sul fronte terminologico, le annotazioni riguardano 3 errori di traduzione di termini (A/Termi/SemUn) e 2 errori di traduzione non necessaria di termini (A/Termi/DNT). Questi dati indicano una presenza limitata di problemi di natura specialistica. Si segnalano infine 12 occorrenze per la categoria F/Coh/Cit, riferibili alla strutturazione incoerente dei riferimenti bibliografici.

Durata delle sessioni di post-editing con Qualitivity

Il *post-editing* è stato condotto in sei sessioni della durata rispettiva di 03:10:56, 02:51:41, 00:55:15, 03:22:51, 03:06:13, 00:14:42. Il totale ammonta a 13:41:38, pause escluse.

Report di Post-Edit Compare

Il report analizza 296 segmenti, per un totale di circa 7.217 parole. Tutti i segmenti sono generati dal sistema di TA, senza contributi provenienti dalla memoria di traduzione, a conferma che il testo di arrivo sottoposto a *post-editing* si basa esclusivamente sui risultati della TA.

Fasce di PEMP	Segmenti	Parole	Percentuali
100%	62	880	12,19%
95-99%	39	1.160	16,07%

85-94%	74	2.509	34,77%
75-84%	23	830	11,50%
50-74%	20	539	7,47%
New (0-49%)	78	1.299	18,00%

Tabella 14 (All. 4): Distribuzione delle fasce di PEMp per testo n. 7 del progetto n. 3

La distribuzione delle parole rivela una qualità complessivamente eterogenea degli interventi. Una parte relativamente limitata del testo, cioè il 12,19%, è stata accettata senza modifiche; mentre segmenti che hanno richiesto interventi minimi, cioè la fascia 95-99%, rappresentano il 16,07% del totale delle parole. La fascia 85-94% è la più consistente, coprendo il 34,77% del testo, e indica che oltre un terzo del testo è stato sottoposto a correzioni di lieve entità. I segmenti che hanno richiesto correzioni medie e medio-alte si collocano nella fascia 75-84%, che registra l'11,50%, e in quella 50-74%, che copre il 7,47%, indicando una diffusione limitata. La categoria *New* (0-49%) rappresenta il 18,00% delle parole. Si tratta di una parte di testo non trascurabile, poco meno di un quinto del testo, che è stata sottoposta a rielaborazioni complete o quasi complete.

In generale, l'output della TA ha fornito una base utile ma non uniforme: una parte ridotta è stata riutilizzata senza modifiche, altre hanno richiesto interventi di varia entità. La constatazione nel testo di segmenti riutilizzati senza modifiche e di una quota significativa per la categoria *New* (0-49%) mette in luce quanto il contributo del *post-editor* sia determinante per assicurare la qualità desiderata. L'attività di *post-editing* quindi si è sviluppata su più livelli, comprendendo correzione più superficiali e riformulazioni più o meno estese, con una predominanza degli interventi di lieve entità.

Testo n. 8 (6.120 tokens)

Stang, Alexandra. 2022. «Theoretische Reflexionen und Konturen einer relational orientierten Wissenskommunikation. Überlegungen für eine künftige Forschungsagenda». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 21 (35): 47–64. <https://doi.org/10.24403/jp.1295924>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	118
F/Sty/Disfl	78
F/Lex/LexCho	71
A/MisTra/SemUn	57
F/Coh/Cit	33
A/Untra	25

A/Omis	22
A/Mech/Punct	13
A/ForLa	10
A/DNT	9
A/MisOrd/WO	9
A/Termi/SemUn	9
A/MisTra/WSD	9
F/Sty/Reg	7
A/SouEr	7
G/Orth/Capit	7
A/MisTra/POS	6
G/Orth/WhiSpa	4
F/Lex/Fue	4
A/MisTra/Part	4
F/SLP/Nomin	4
G/WF/Agree	4
F/Coh/CoRef	4
G/Orth/Spel	3
G/Orth/Punct	2
A/MWE/Komp	2
A/Mech/CharFor	2
A/Add	2
A/MisTra/WroPr	2
F/Sty/Idiom	2
A/MWE/Ell	2
A/Termi/DNT	2
F/De/Layo	2
F/De/CharForm	2
A/MisExpl	2
A/Oth	1

Tabella 15 (All. 4): Analisi quantitativa degli errori per il testo n. 8 del progetto n. 3

Nel corso dell'analisi si riscontrano 325 errori di *Fluency*, 205 di *Accuracy* e 20 di *Grammar*. Il totale degli errori annotati ammonta a 550. La maggior parte riguarda la *Fluency* e supera la metà del totale, principalmente a causa della categoria F/Coh/InPu, che conta 118 occorrenze di uso incoerente delle virgolette e dell'apostrofo. In ordine decrescente per numero di occorrenze, seguono le seguenti categorie di errori: F/Sty/Disfl per le problematiche di scorrevolezza della lingua di arrivo (78 errori); F/Lex/LexCho per la scelta lessicale errata in riferimento all'ambito e al genere testuale trattati (71 errori) e A/MisTra/SemUn per i casi di mancata corrispondenza tra il testo di partenza e quello di arrivo (57 errori). Particolarmente rilevante è la categoria F/Coh/Cit che registra 33 errori e riguarda la strutturazione incoerente dei riferimenti bibliografici delle citazioni nel testo. Per quanto concerne gli errori stilistici del

linguaggio scientifico, si registrano 7 errori di registro inadeguato (F/Sty/Reg), 4 errori di traduzione di *Fügungen* e 4 errori di mancata realizzazione della nominalizzazione (F/SLP/Nomin). In ambito terminologico, si contano 9 errori di traduzione errata (A/Termi/SemUn) e 2 errori di traduzione non necessaria (A/Termi/DNT).

Durata delle sessioni di post-editing con Qualityity

Il *post-editing* si è svolto in cinque sessioni della durata rispettiva di 01:35:58, 04:30:00, 01:40:31, 04:15:43, 00:21:05, per un totale complessivo di 12:23:17.

Report di Post-Edit Compare

Il report di Post-Edit Compare analizza 223 segmenti, per un totale di circa 4.996 parole. La maggior parte del testo, 219 segmenti, è generata dal sistema di TA DeepL, mentre 4 segmenti vengono introdotti manualmente come *New*. Non emergono proposte della memoria di traduzione, mettendo in luce che il materiale esaminato deriva interamente dall'output automatico.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	39	592	11,85%
95-99%	39	1.066	21,34%
85-94%	44	1.480	26,42%
75-84%	18	510	10,21%
50-74%	26	675	13,51%
New (0-49%)	57	833	16,67%

Tabella 16 (All. 4): Distribuzione delle fasce di PEMp per testo n. 8 del progetto n. 3

La tabella 67 mostra una distribuzione delle fasce di PEMp piuttosto varia. La fascia 100% copre l'11,85% delle parole, indicando che solo una porzione limitata del testo è stata accettata senza modifiche. Segmenti con correzioni minime, rientranti nella fascia 95-99%, rappresentano il 21,34% delle parole, suggerendo che oltre un quinto del testo ha subito delle micro-correzioni mirate. Diversamente, la fascia 85-94% costituisce la porzione più consistente, con 26,42% delle parole. Questo dato risulta particolarmente significativo poiché indica che più di un quarto del testo è stato modificato con un'intensità media. Le fasce 75-84% e 50-74% raccolgono rispettivamente il 10,21% e il 13,51% delle parole, rappresentando correzioni più approfondite su una porzione di segmenti limitata ma non trascurabile. Infine, la categoria *New* (0-49%) incide per il 16,67% delle parole, a indicare la necessità di riscritture totali o quasi totali su porzioni non marginali del testo.

Nel complesso, l'output della TA ha fornito una base utile, ma non uniforme: alcune parti sono risultate utilizzabili così come prodotte automaticamente, mentre altre hanno richiesto interventi più o meno estesi. La percentuale di segmenti 100% è piuttosto basse e la restante parte di

segmenti è stata interessata da modifiche di varia entità, con una prevalenza di interventi minimi e lievi. Le percentuali registrate per le fasce di interventi medi, medio-alti e integrali suggeriscono la centralità dell'intervento umano in fase di *post-editing* per raggiungere un testo con una qualità pubblicabile.

Testo n. 9 (10.330 tokens)

Seyferth-Zapf, Maria, e Silke Grafe. 2022. «Gemeinsame Gestaltung medialer Beiträge als Ziel interkultureller Begegnung». *Interculture journal: Online-Zeitschrift für interkulturelle Studien* 21 (35): 87–108. <https://doi.org/10.24403/jp.1295945>.

Analisi quantitativa degli errori

Categoria di errore	Numero di annotazioni
F/Coh/InPu	251
F/Sty/Disfl	161
A/MisTra/SemUn	122
F/Lex/LexCho	101
F/De/CharFor	70
A/MisTra/Gend	67
A/Untra	49
A/MisTra/Part	46
A/MisTra/WSD	41
A/Omis	33
A/DNT	32
F/SLP/Nomin	20
G/Orth/Capit	18
A/MisTra/POS	17
A/MisOrd/WO	15
A/Mech/Punct	15
F/Coh/CoRef	13
G/Orth/Punct	12
F/Coh/Cit	11
F/Lex/Fue	9
A/MisTra/WroPr	9
G/WF/Agree	8
F/Sty/Reg	7
G/Orth/Spel	7
F/De/Layo	5
A/Add	5
A/ForLa	4
A/Termi/SemUn	4
A/MWE/Ell	3
A/Oth	1

G/Orth/WhiSpa	1
A/MisExpl	1
A/MisOrd/MW	1
A/MWE/Komp	1
A/Termi/DNT	1

Tabella 17 (All. 4): Analisi quantitativa degli errori per il testo n. 9 del progetto n. 3

Nel complesso, si rilevano 648 errori di *Fluency*, 467 di *Accuracy* e 46 di *Grammar*, per un totale di 1.161 annotazioni. La *Fluency* si conferma la macrocategoria più problematica, con oltre la metà delle occorrenze totali: da sola, la sottocategoria F/Coh/InPu, che segnala i casi di uso incoerente di virgolette e apostrofo, raggiunge 251 errori. Seguono per rilevanza gli errori riferibili alla mancata corrispondenza semantica tra il testo di arrivo e quello di partenza (A/MisTra/SemUn, 161 occorrenze), alla scelta lessicale inadeguata all'ambito o al genere testuale presi in esame (F/Lex/LexCoh, 122 occorrenze) e alla scorrevolezza della lingua di arrivo (F/Sty/Disfl, 101 occorrenze). Un numero consistente di errori si registra anche per le categorie F/De/CharFor con 70 occorrenze e A/MisTra/Gend con 67 occorrenze. La prima categoria è legata alla formattazione del testo di arrivo, in questo caso in particolare all'uso nel testo di arrivo di una parola in tedesco e alla necessità di riportarla, in quanto parola straniera, in corsivo. La seconda categoria è riconducibile all'uso del linguaggio inclusivo e, quindi, alla mancata resa nel testo di arrivo della forma femminile accanto a quella maschile, come nel caso della formulazione “Schülerinnen und Schüler”, che nella TA risulta genericamente resa come “studenti”. Altre occorrenze considerevoli si verificano nelle seguenti categorie: A/Untra per i segmenti non tradotti (49 errori); A/MisTra/Part per gli errori riguardanti il trasferimento semantico parziale (46 errori); A/MisTra/WSD per gli errori di scelta semantica in riferimento a parole polisemiche (41 errori); A/Omis per i segmenti omessi (33 errori) e A/DNT per i segmenti da non tradurre (32 errori). All'interno della categoria A/MisTra/Part, 28 occorrenze segnalano un *gender bias*, ovvero l'uso esclusivo del maschile in contesti in cui era richiesta la forma femminile. Gli errori stilistici legati al linguaggio scientifico comprendono 20 errori di nominalizzazione non realizzata (F/SLP/Nomin), 9 errori di traduzione di *Fügungen* (F/Lex/Fue) e 7 errori di registro inappropriato (F/Sty/Reg). Per quanto riguarda invece la terminologia, si annotano 4 errori di traduzione errata di termini (A/Termi/SemUn) e un errore di traduzione non necessaria di termini (A/Termi/DNT). Infine, si individuano 11 errori per la categoria F/Coh/Cit che assume rilievo nel contesto dell'articolo scientifico, poiché riguarda riferimenti bibliografici riportati in maniera incoerente.

Durata delle sessioni di post-editing con Qualitivity

Il *post-editing* si è svolto in cinque sessioni della durata di 03:28:10, 03:57:57, 00:22:24, 03:43:32, 04:53:02, per un totale complessivo di 16:25:05.

Report di Post-Edit Compare

Il presente report prende in esame un testo di 421 segmenti, pari a 8.472 parole complessive, interamente generate dal sistema di TA. Non si rilevano proposte provenienti dalla memoria, pertanto l'analisi si basa esclusivamente sull'output della TA.

Fasce di PEMp	Segmenti	Parole	Percentuali
100%	60	815	9,62%
95-99%	50	1.060	12,51%
85-94%	102	2.507	29,59%
75-84%	60	1.568	18,51%
50-74%	41	937	11,06%
New (0-49%)	108	1.585	18,71%

Tabella 18 (All. 4): Distribuzione delle fasce di PEMp per testo n. 9 del progetto n. 3

La distribuzione dei segmenti per fasce di PEMp mostra un quadro articolato. La fascia 100% rappresenta soltanto il 9,62% del totale, suggerendo che solo una quota limitata di testo è stata accettata senza modifiche. Nella fascia 95-99% rientrano il 12,51% delle parole che hanno richiesto interventi minimi. La quota più consistente si registra nella fascia 85-94% con il 29,59% delle parole, indicando che quasi un terzo è stato sottoposto a interventi di lieve entità. Le fasce che segnalano interventi di intensità media e medio-alta presentano i seguenti dati: la categoria 75-84% raccoglie il 18,51% del testo, mentre la fascia 50-74% incide per l'11,06%. Si tratta in entrambi i casi di risultati non trascurabili. Infine, i segmenti etichettati come *New* (0-49%) costituiscono una porzione significativa, pari al 18,71% delle parole, attestando che circa un quinto del testo ha richiesto rielaborazioni integrali o ampiamente invasive.

Nel complesso, emerge il profilo di un output della TA solo parzialmente sfruttabile: la porzione di testo che può essere già utilizzata o che ha richiesto degli interventi minimi risulta limitata (circa il 20%). La restante parte ha richiesto interventi di varia entità, con una prevalenza degli interventi di lieve portata. I risultati evidenziano quindi l'importanza del ruolo del *post-editor* per garantire la qualità del prodotto finale con un *effort* richiesto di varia portata.

Riflessione conclusiva

Complessivamente, le tabelle quantitative dei singoli testi mostrano che la macrocategoria di errori *Fluency* registra sistematicamente il numero più elevato di occorrenze. Segue la macrocategoria *Accuracy* e, infine, *Grammar*, caratterizzata dal numero minore di errori. Ciò suggerisce che, dal

punto di vista quantitativo, le problematiche maggiormente ricorrenti interessano la scorrevolezza e la naturalezza della lingua di arrivo, aspetti centrali ai fini della pubblicabilità del testo scientifico. Per un'analisi qualitativa e maggiormente dettagliata delle tipologie di errore si rimanda al capitolo 6, paragrafo 6.3.

Dall'esame dei dati relativi ai nove testi del progetto emerge una correlazione tendenziale tra la lunghezza dei testi e la durata complessiva delle sessioni di *post-editing*. Sebbene i testi più estesi richiedano generalmente tempi maggiori, il caso del testo n. 5 (8.706 *tokens*) risulta emblematico: a fronte di un'estensione intermedia, registra la durata più consistente (17:56:41), superiore anche a quella di testi più estesi. Ciò conferma che la qualità dell'output generato dalla TA incide sulla durata del *post-editing* più del solo volume del testo.

Analogamente, il confronto tra numero di errori e tempo di correzione mostra una tendenza complessiva coerente ma non priva di eccezioni. Nella maggior parte dei casi, un numero più elevato di errori corrisponde a tempi di *post-editing* più lunghi; tuttavia, alcuni testi si discostano da questa regola generale. Il testo n. 8, ad esempio, pur essendo il meno esteso (6.120 *tokens*) e contenendo un numero relativamente ridotto di errori (534), presenta una durata di correzione superiore a quella di testi di dimensioni simili. Inoltre, dal rapporto tra numero di *tokens* e tempo di *post-editing* si può stimare una produttività media compresa tra circa 520 e 670 *tokens*/ora, con scostamenti significativi tra i vari testi.

In riferimento ai risultati dei report di Post-Edit Compare, la distribuzione delle fasce di *post-editing* risulta complessivamente eterogenea, senza evidenziare pattern stabili o ricorrenti tra i testi. La fascia 85-94%, che indica le correzioni lievi, rappresenta la quota più consistente, con valori compresi tra il 26% e il 35%, fatta eccezione per il testo n. 2 e n. 4. Per la fascia 100% (segmenti non modificati) si osservano due tendenze: quattro testi presentano valori tra il 9-12%, mentre altri quattro tra il 19-24%. Un solo testo registra il 29,54 discostandosi visibilmente dagli altri risultati. Anche nella categoria *New* (0-49%, correzioni integrali o quasi) emergono due gruppi distinti: cinque testi presentano valori tra il 15-19%; tre testi tra il 20-24%. Anche in questo caso si osserva un solo testo non riconducibile alle due fasce individuate che registra il 32,10%. Le fasce 95-99% (correzioni minime), 75-84% (correzioni medie) e 50-74% (correzioni medio-alte) presentano valori ampiamente disomogenei: per la fascia 95-99% i risultati vanno da 11% a 22%; per 75-84% da 5% a 19% e per 50-74% da 5 a 14%.

Nel complesso, questa distribuzione frammentata evidenzia l'assenza di regolarità nella qualità dell'output della TA, che tende a variare sensibilmente non solo da un testo all'altro, ma anche tra segmenti dello stesso documento. Tale variabilità incide direttamente sull'*effort* del *post-editor*, costretto ad alternare fasi di correzione minima a interventi più profondi, con un conseguente

aumento del carico cognitivo. Questi risultati confermano quindi la scarsa affidabilità della traduzione automatica per porzioni significative di testo e sottolineano la necessità imprescindibile dell'intervento umano per garantire coerenza, leggibilità e adeguatezza complessiva del prodotto finale.