Corpus studies

In brief

ITA *corpus* **SPA** *corpus*

other names

Research in this area has been commonly referred to as Corpus-based Translation Studies (CBTS, CTS), Corpus-based Interpreting Studies (CIS).

abstract

Corpus-based research into translation and interpreting, taken here to subsume *Corpus-based Translation Studies* and *Corpus-based Interpreting Studies*, is a branch of Translation Studies in which corpus analysis is used as a major paradigm and research methodology for the analysis of translated and interpreted language and what makes them distinct from non-mediated language. This entry covers topics that have been at the core of CTS/CIS research, such as translation universals, translator/interpreter style and other quantitative studies into translation and interpreting.

record

Sara Castagnoli & Marta Kajzer-Wietrzny

2024

Castagnoli, Sara & Marta Kajzer-Wietrzny. 2024. "Corpus studies" @ *ENTI (Encyclopedia of translation and interpreting).* AIETI.

https://doi.org/10.5281/zenodo.10938945

https://www.aieti.eu/enti/corpus_ENG

Entry

ITA *corpus* **SPA** *corpus*

contents

Introduction

Corpus-based research into translation initiated in the 1990s, at a time when corpus linguistics had started to establish itself as a mainstream methodology for the study of language. Corpus linguists argued that language should be studied in genuine instances of use, i.e. in naturally occurring texts, rather than through invented examples of what is theoretically possible in a language. The possibility to create *corpora* – that is, large collections of authentic texts, selected according to explicit design criteria to be representative of a specific language variety, and stored in electronic form so as to be searched through dedicated software applications – and to use them to detect and analyse linguistic patterns thus began to be considered as a fruitful methodology for both descriptive/theoretical research about language and for applied purposes.

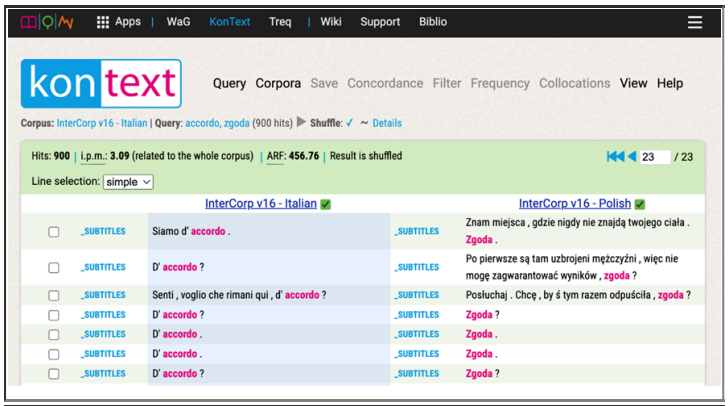
The first scholar to suggest that the availability of corpora and of corpus-driven methodologies could also provide valuable insights in Translation Studies (TS) was arguably Baker (1993), who highlighted their potential for identifying recurring features of translated texts which would help elucidate the nature of translation as a mediated – and thus distinct – communicative event. Baker's seminal paper followed on other prominent translation scholars who had started to plead for the recognition of translated texts as linguistic products in their own right rather than as flawed reconstructions of the corresponding source texts. According to Toury (1981), for example, translation was to be considered as a language form of its own, whose general laws and regularities deserved to be the focus of translation studies scholars:

> [i]t is highly probable that there are "dialectical" differences between texts originally composed in TL and translations into it [...]. As a matter of fact, [...] "deviations" of translations both with respect to ST- or SL-based models of "functional equivalence" and to TL-normality are the major justification for a distinct discipline of translation studies which […] will focus on translational phenomena per se. (Toury 1981: 16)

The end of the 20th century thus marked a dramatic change in the focus of TS: from source-oriented, equivalence-focused and largely prescriptive approaches to translation, to target-oriented

and descriptive approaches aimed at identifying and accounting for regularities characterising translated language. The terms *translationese* and, later, *interpretese* started to be used to denote the supposed distinctiveness of translated and interpreted language as mediated language varieties.

Corpora began to be considered as a possible response to the emerging call for greater empiricism in translation studies, allowing researchers to overcome reliance on sparse, anecdotal evidence. Building on Baker, Shlesinger (2009) further suggested that the benefits of the corpus-based methodology could be extended to the study of interpreting, as well as to intermodal comparisons revealing differences between translation and interpreting as distinct mediation modes.



Parallel concordance (Italian-Polish) from the Intercorp corpus available at korpus.cz. The concordance can be reproduced here.

Two major corpus typologies are used in corpus translation studies (CTS) and corpus interpreting studies (CIS), namely *monolingual comparable corpora* and *(bilingual) parallel corpora*. Monolingual comparable corpora (MCC) are collections of original (i.e. non-translated) and translated texts in the same language, selected so as to be comparable/similar in terms of register, topic, publication date etc. Accordingly, and largely following the approach proposed by Baker, MCC have been considered as privileged tools to observe how translated language differs from non-translated language, comparing the two relevant subcorpora on the basis of indicators and quantitative measures – such as type-token ratio, lexical density, frequency of specific words, word classes or collocations – that can be automatically extracted from texts. Within parallel corpora, on the other hand, source texts in one language are collected together with their translations into a target language. Corresponding source-text and target-text units (generally at the level of sentences) are usually *aligned*, i.e. they are automatically identified and linked in order to enable *bilingual concordancing*: this means that searching for an item in one language allows researchers to retrieve all sentences containing that item, along with their corresponding sentences in the other language, to observe possible equivalent items. As two languages are involved, parallel corpus investigations are more laborious than MCC investigations, since the data extracted from source and target texts need to be analysed also in the light of systemic and pragmalinguistic differences between languages (which can only be observed through contrastive analysis), and translation shifts – indicating conscious or unconscious translator strategies – need to be investigated using a more qualitative approach.

As for language corpora in general, translation-oriented corpora are usually enriched with different levels of annotation in order to enable scholars to perform fine-grained and targeted analyses. Linguistic information (mainly parts-of-speech, lemmas, syntactic parsing – see Computational linguistics) and structural information (e.g. sentence boundaries) is usually added through automatic processing to enable more sophisticated queries than for simple word forms, as well as the sentence-alignment of parallel corpora. An additional layer of annotation, whose importance has increasingly been emphasised in recent years, comprises metadata about the translator (e.g. their

professional expertise/experience, as well as their linguistic, sociological and educational background) and the translation task (e.g. text type, translation brief, use of translation software, editorial intervention), that is, information about extra-linguistic, contextual factors that can contribute to shaping translated language.

back to top

¶ Corpus studies about translation

Contemporary CTS are characterised by several lines of inquiry. The search for regularities that has characterised descriptive TS in general and CTS in particular since their inception is today accompanied by an increased attention to variation associated to either individual translators or contextual factors. Consequently, traditional corpus resources and methods are increasingly accompanied by new corpus designs and research methods, as shown in the following sections.

Translation universals

Early CTS research was primarily dedicated to finding evidence of the existence of *universal features of translation* (or *translation universals*). The concept had been introduced in Baker's (1993) programmatic paper to refer to "features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems" – in other words, irrespective of the source and target languages involved – "as a product of constraints which are inherent in the translation process itself" (Baker 1993: 243-246). These included, among others that have been given less consideration in subsequent studies: a tendency towards *explicitation,* defined as "a marked rise in the level of explicitness compared to specific source texts and to original texts in general" (Baker 1993: 243), by far the most investigated and the most convincingly supported potential translation universal; a tendency to lexical and syntactic *simplification*, and to use more conventional lexis and syntax than source and comparable texts, known as *normalisation* or *conservatism*. Even though Baker's proposal prioritised research based on monolingual comparable corpora, contrasting translations with non-translated texts in the same language, the existence of two distinct research foci was pointed out by Chesterman (2004: 39), who suggested that a basic distinction was to be made between potential universal differences between translations and their source texts, or *S-universals* ("i.e. characteristics of the way in which translators process the source text"), and between translations and comparable non-translated texts, or *T-universals* ("i.e. characteristics of the way translators use the target language"). The distinction is important for two main reasons. On a theoretical level, because the two types of analyses concern two key but separate aspects of research on translation, namely the equivalence relation with the source text, and the relation of textual fit with comparable non-translated target language texts. On a more methodological level, because they require different corpus resources: S-universals can be investigated through parallel corpora, while researching T-universals requires a comparison between translated texts and comparable non-translated texts in the target language, which can be obtained with monolingual comparable corpora.

Initial empirical research was based on the search for a relatively limited number of textual patterns considered to be indicative of the existence of translation universals, generally within limited datasets. As regards explicitation, for example, it was explored as a S-universal by looking at the

frequency of linking adverbials in source texts (ST) and target texts (TT), taken as an indicator of the relative explicitness of clausal relations; shifts from non-finite to finite clauses; shifts in lexical cohesion, for example through the lexicalisation of ST pronominal forms. Bi-directional parallel corpora collecting L1>L2 as well as L2>L1 translations were sometimes used to factor in systemic differences and the directionality of translation as causal explanations for some of the observed patterns. From a MCC perspective, besides considering the above indicators in translated versus original comparisons, the higher frequency of optional items (e.g. the English complementiser *that*, or subject pronouns in pro-drop languages) in translated than in original language was often taken as proof of the greater explicitness of the former. Simplification, on the other hand, was often investigated as a T-universal by taking into account measures like lexical variety and lexical density, or sentence length, which were observed to be lower in translated language than in original language (see e.g. Laviosa 1998). Research on normalisation also followed a twofold perspective. From the parallel perspective, some studies suggested that target texts could be seen as “sanitized” or standardised versions of the corresponding STs, toning down creative usage and displaying more conventional vocabulary and grammar. In a MCC perspective, on the other hand, a number of studies have shown that features such as contracted forms and optional *that* omission (as regards English language texts), or anglicisms (in languages other than English - see e.g. Bernardini & Ferraresi (2011) for Italian), are less frequent in translated than in original texts in the same language, which would suggest that translators make more conservative choices than authors of original language.

In some cases, conflicting hypotheses were put forward: for example, the claim that translated texts would tend to conform to, or even exaggerate, typical target-language patterns was somehow at odds with the *unique item hypothesis*, according to which typical features of the target language lacking straightforward linguistic counterparts in the source language would tend to be under-represented in translation (e.g. Tirkkonen-Condit 2004).

Theoretical and methodological reservations about the tenability of these concepts, their operationalisation and their universal status soon emerged. Most notably, serious reservations were expressed regarding the substantial neglect of the influence of the source text that Toury (1995), on the other hand, had theorised through his *law of interference*. Notwithstanding criticisms (possibly caused by a partial interpretation of Baker’s research agenda, as remarked De Sutter & Lefer 2020: 2), some of these concepts have remained at the core of the CTS agenda, and continue being investigated through sophisticated corpus types and analytical tools which take into due consideration the importance of the ST/SL as well as of variables other than the translated vs. non-translated dichotomy.

In more recent years it has been suggested that some of the observed phenomena are not only attributable to the difference between translated and non-translated texts, but also depend and co-vary with other factors such as text type, source language, translator expertise, translation mode and translation method. The second decade of the 21st century has thus witnessed the birth and rise of more sophisticated corpus designs and research methods (see e.g. contributions in Oakes & Ji (2012)), and most notably of multifactorial investigations taking into account a larger number of variables through multivariate analysis, often based on a combination of MCC and parallel corpus data. Delaere, De Sutter & Plevoets (2012), for instance, rely on a corpus of non-translated Belgian Dutch and Belgian Dutch translations from English and French to investigate whether the level of

standardisation – measured as the frequency of occurrence of non-standard Belgian Dutch lexical items for which a standard variant exists – differs in the three corpus components as well as across the six different text types included in each component. The results of their profile-based correspondence analysis confirm a tendency for translations to use more standard language than non-translated texts, especially within text types characterised by a lot of editorial control (fiction, non-fiction and journalistic texts) compared to the less edited text types (administrative texts and external communication). As regards the effect of the source language on translational products, Cappelle & Loock (2017) investigate whether there is a difference in the usage of phrasal verbs in English translations from Romance and Germanic languages. The study shows that the typological similarity or difference between SL (source language) and TL (target language) has a significant effect on translation: translations from Romance languages contain fewer phrasal verbs than translations into English from Germanic languages, which do not differ noticeably from non-translated English in the frequency of phrasal verbs. The study thus questions the validity of normalisation as a translation universal, and rather supports the claim that interference is the most remarkable and inherent characteristic distinguishing translated from non-translated language.

In addition to this surge in multifactorial investigations, CTS is also becoming more and more multi-methodological, as corpus data is increasingly triangulated with information obtained through translation process research methods – such as keylogging, eye-tracking, screen recording – and cognitive translation studies, which can provide insights about cognitive mechanisms that can supposedly influence and explain some translation features. (see e.g. Halverson 2017).

Translator style and variation

A second research strand in CTS revolves around the comparison of the behaviour of different translators in order to investigate translator style or the issue of variation in translation. This is increasingly carried out using multiple translation corpora, i.e. parallel corpora which contain several concurrent translations into the same target language for each source text, and thus allow researchers to analyse alternative solutions for given source-text items.

Translator style and individual variation

Translator style is conceived in terms of features of translated texts that, besides not being related to the style of the source text or to target language requirements, are “recognizable across a range of translations by the same translator” and “distinguish that translator’s work from that of other translators” (Saldanha 2011: 31). It represents a recent research strand as in the era of source-oriented prescriptive TS there was little interest in studying the style of individual translators: in fact, translation was seen as a derivative act and, consequently, translators were supposed to reproduce as closely as possible the style of the source text rather than imposing their own on the translated text.

Literary texts have often and understandably been considered an ideal testbed for CTS about translator style, also due to the fact that very few other genres tend to get translated multiple times and allow researchers to analyse comparable translations into the same target language. At the beginning of the 21st century, research aimed at uncovering evidence of the translator’s voice and at comparing strategies adopted by different translators was carried out e.g. by Winters (2009), who

analysed the use of loan words, code switches, speech-act report verbs and modal particles in two German translations of Fitzgerald's *The Beautiful and Damned*. Subsequent quantitative studies, such as Li, Zhang & Liu (2011), went beyond the analysis of selected ST features and, following a target-oriented approach, tried to compare translations considering corpus-driven information such as type/token ratio and sentence length, keywords and key clusters. Statistical analyses carried out by Mikhailov & Villikka (2001) on a multiple translation corpus of Russian-to-Finnish translations ruled out the possibility to identify translators through indicators usually exploited in authorship attribution tasks like vocabulary richness, frequent and favourite words; however, the authors managed to identify other indicators shaping translator profile, such as the use of modals and conjunctions, splitting or joining of sentences, or shortening or expanding of the text. Clause building and clause reduction were also the focus of studies based on the English–Norwegian Multiple Translation Corpus which found translators to differ in their tendency to preserve or modify the syntactic structure of the original.

Somewhat related to translator style, *multiple translation corpora* can be used to investigate variation connected to translator competence: the term subsumes here not only the individual translator's expertise (basically, professional vs. trainee), but also their cognitive, linguistic, sociological and educational background, which determines their use of language and their approach to translation. Multiple translation corpora containing professional and student translations, often contrasted to non-translated comparable texts, have been investigated using advanced statistical methods to find correlations between specific translation features and different levels of expertise (e.g. Lapshinova-Koltunski 2022). In a recent study, Popovic, Lapshinova-Koltunski & Koponen (2023) introduced a third element of comparison and interestingly contrasted professional and student translations with machine translation output. Comparing the three translation varieties in terms of lexical and grammatical variety, sentence length, and part-of-speech (POS) tags and POS-trigrams frequencies, the authors noticed that student translations are sometimes more similar to machine translations than to professional translations, especially regarding their closeness to the structure of the source text.

Learner translator corpora

As the number of texts that authentically get retranslated multiple times is extremely limited, creating multiple translation corpora can be an arduous task, and scholars have to either commission alternative translations of the same source text or collect them in relevant experimental settings (as in the case of the Translation Process Research Database). A convenient alternative is including student translations, either in contrast with professional translation (see above) or as an independent object of study, most notably in the framework of learner translation corpus research (Castagnoli 2022). Research on learner translation can be seen as standing at the crossroads between corpus translation studies and learner corpus research, with translation students representing a special type of learners whose use of language in translation is the focus of analysis. While *learner translation corpora* (LTC) can be used to investigate how typical CTS questions relate to learner translation, error-annotated LTC can help scholars determine, for instance, which are the most typical difficulties for trainees at specific levels of instruction (e.g. beginners vs. advanced students), with respect to source texts belonging to different text types, or if the frequency of specific error types decreases after targeted teaching. The descriptive potential of LTC is therefore associated with a strong

pedagogical potential, as insights from LTC research can naturally inform translation teaching (e.g. Kunilovskaya, Ilyushchenya, Morgoun *et al.* 2022).

Regularities and variation

Another promising application of multiple translation corpora is the study of the interplay between regularities and variation in translation, namely the analysis of similarities and differences that occur between alternative renditions of the same source text. In fact, even though alternative translations of a ST inevitably display a high degree of similarity, because of the existence of an anterior text whose meaning and wording need to be re-coded, variation is also expected to occur as translators – like all language users – have a range of more or less likely linguistic options at their disposal to express meaning. Multiple translation corpora thus allow scholars to identify phenomena representing real translation regularities – i.e. patterns observed in the behaviour of different translators vis-à-vis the same ST – and to investigate (what can determine) translation variability.

A line of investigation that appears particularly insightful associates invariance across alternative translations with the possibility of a literal translation of the ST (e.g. Carl & Schaeffer 2017), which would lead translators not to override the priming effect of the ST. On the contrary, variation would mainly occur with respect to items for which a literal translation is not possible, as well as for items with the highest interpretative and evaluative potential, such as words expressing attitude and ideology, abstract nouns and idiomatic expressions, which require a higher processing effort. From this perspective, variation across alternative translations might be indicative of translation difficulty, and some studies have tried to combine corpus data and process-based insights to investigate the possible relationship between variation and the translation difficulty of specific ST items, as signalled by indicators of increased cognitive effort in process data (e.g. Dragsted 2012).

Regularities and/or variability could be associated with characteristics that are shared by – or that differentiate – translators, such as their level of expertise (see above), or their interaction with technology. Some recent studies, for example, build on corpora which include both human translations and post-edited translations alongside MT output, reflecting the great transformation witnessed by the translation market in the last decade (see e.g. Czulo, Hansen-Schirra & Nitzke 2017 on terminological variation across the three types of translations).

back to top

¶ Corpus studies about interpreting

Interpretation corpora as a challenge

Just as speech is very different from writing, interpreting is different from translation and compiling interpreting corpora is different from compiling translation corpora. Interpretation involves oral discourse production which is far less organized and smooth than written discourse. Interpretation, as speech, is very much embedded in the context in which it was produced and is rarely received without it thus interpreting corpora usually are enriched with complex annotations.

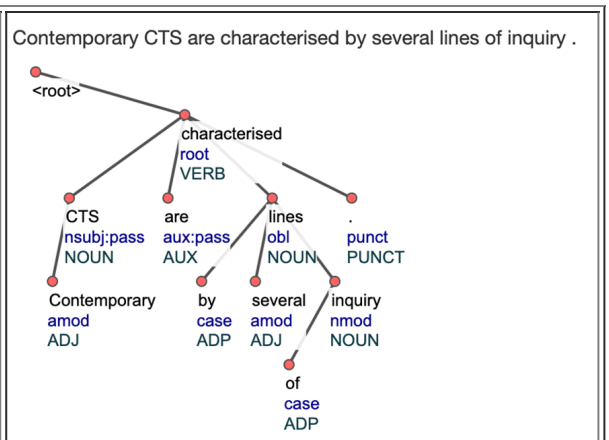
Like spoken corpora, interpreting corpora may contain an orthographic transcript of what was said but increasingly more often transcripts are available together with annotations that extend to a rich

set of paralinguistic features with a great level of details such as speech rate, false starts, length of pauses etc. Transcription conventions used in the compilation process have to be carefully designed to meet the required research goals. One of the first transcription conventions for an interpreting corpus has been designed by the creators of the EPIC corpus and these usually serve as a starting point for other researchers (e.g. EPIC-UDS or EPTIC).

Spoken and interpreted discourse also poses some limitations for automatic taggers and linguists analysing interpretations frequently face challenges like establishing the sentence, or rather, utterance boundary, which, on the other hand, is inextricably linked with annotating the text for Parts of Speech and Universal Dependencies. With the boundaries of an oral utterance being so fluid, the alignment of interpretations and their source speeches becomes more laborious than in the case of written translations. Unlike in the case of written translations, automatic alignment of spoken texts is, for now, very unreliable. Nevertheless today, most of the interpreting corpora are both comparable and parallel and aligned at sentence level, with a few also including alignment with corresponding translated texts and original videos (EPTIC), or even sound-to-text alignment at word level (PINC).

Id	Form	Lemma	UPosTag	XPosTag	Feats	Head	DepRel	Deps	Misc
# generator = UDPipe 2, https://lindat.mff.cuni.cz/services/udpipe									
# udpipe_model = english-ewt-ud-2.12-230717									
# udpipe_model_licence = CC BY-NC-SA									
# newdoc									
# newpar									
# sent_id = 1									
# text = Contemporary CTS are characterised by several lines of inquiry.									
1	Contemporary	contemporary	ADJ	JJ	Degree=Pos	2	amod	_	TokenRange=0:12
2	CTS	ct	NOUN	NNS	Number=Plur	4	nsubj:pass	_	TokenRange=13:16
3	are	be	AUX	VBP	Mood=Ind|Number=Plur|Person=3|Tense=Pres|VerbForm=Fin	4	aux:pass	_	TokenRange=17:20
4	characterised	characterise	VERB	VBN	Tense=Past|VerbForm=Part|Voice=Pass	0	root	_	TokenRange=21:34
5	by	by	ADP	IN	_	7	case	_	TokenRange=35:37
6	several	several	ADJ	JJ	Degree=Pos	7	amod	_	TokenRange=38:45
7	lines	line	NOUN	NNS	Number=Plur	4	obl	_	TokenRange=46:51
8	of	of	ADP	IN	_	9	case	_	TokenRange=52:54
9	inquiry	inquiry	NOUN	NN	Number=Sing	7	nmod	_	SpaceAfter=No|TokenRange=55:62
10	.	.	PUNCT	.	_	4	punct	_	SpaceAfter=No|TokenRange=62:63

Sentence tagged with UD pipe including such linguistic information as lemmas, universal dependencies, parts of speech and other



Universal dependencies tree generated with UD pipe

The complex compilation process of interpreting corpora affects their size and frequency of creation and usually is a task that calls for a collaboration of a group of researchers. With the advent of automatic transcription methods, more and more interpreting corpora are made but even greater technological advancement will be necessary for interpreting corpora to grow significantly larger. At the moment both the transcription process and the alignment process, as well as annotation require a lot of human supervision and many time-consuming corrections are usually necessary.

Unimodal and inter-modal comparisons

The number of corpus studies on interpreting is constantly growing as automated analysis of large amounts of authentic interpreting data yields valuable results.

Contrasting interpretations to non-interpretations in the same language using comparable corpora can help researchers closely investigate the characteristic features of interpreted language at the level of lexis, grammar, and syntax as well as paralinguistic features, like disfluencies. Another valuable approach both from the perspective of translation and from the perspective of interpreting is

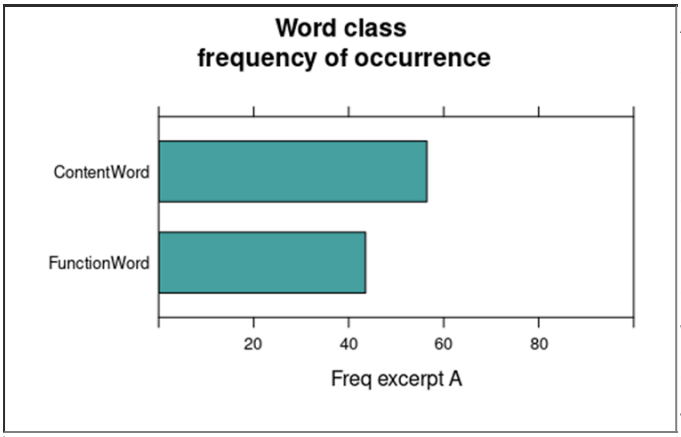
an intermodal analysis of the spoken and written mode of mediation, which can be perceived as a tool for “extrapolating a set of stylistic and pragmatic features” of both varieties (Shlesinger 2009: 237) and allows one to analyse a range of lexico-grammatical features that set the two apart and contributes to drawing a linguistic profile of both interpretese and translationese. As rightly pointed out by Gile (2004: 23), the oral and written mediated variety have so much in common that

> the differences between them can help shed light on each, so that besides the autonomous investigation of their respective features, each step in the investigation of one can contribute valuable input towards investigation of the other.

We will show in subsequent sections that this strand of research has been growing in popularity in recent years and indeed yields promising results.

Lexical patterns

In unimodal analyses, comparable corpora make it possible for scholars to detect patterns that make interpreting distinctive from non-mediated oral speeches. Following translation scholars, interpreting scholars use various measures, particularly suitable for monolingual and comparable investigations to inspect lexical patterns in interpreting, frequently in search of traces of simplification (Laviosa 1998) that would be expected in a linguistic output produced in more cognitively demanding situations. In the oral modality of mediation, parameters typically used to investigate lexical simplification seem to be very dependent on a language pair and directionality. One such measure is lexical density, i.e. the proportion of content words to all words in the text, which indicates how informative a text is (the greater the number of content words per text the higher the information load encoded in it). It seems, however, that this specific parameter behaves differently in interpreting than in translation. Interpretations from various languages into English are either more lexically dense (German, Dutch, French or Russian) or show similar lexical density (Italian and Spanish). The trends are reversed in interpretations from English and Spanish into Italian, just like in interpretations into Russian from English.



Percentage of function words and content words in a text used to estimate lexical density

Another frequently used measure in lexical analyses involves the extent to which language is repetitive. In interpreting, again this parameter also seems to be language-pair dependent. Repetitiveness is frequently measured with list head coverage, which accounts for the proportion of the most frequent words (usually top 100 in the frequency list) in a corpus. Again, while English interpretations from Italian exhibit more repetitiveness compared to original English utterances, just like interpretations from Spanish and German, thus supporting the simplification hypothesis, the same cannot be said about interpretations into Italian from English or into English from French, which show the opposite tendencies (cf. Russo, Bendazzoli & Sandrelli 2006 for detailed results on lexical patterns interpreting). Factors such as the mode of delivery of the original speaker also

appear to influence the tendency towards repetitiveness and other lexical patterns in interpreting (Kajzer-Wietrzny 2015). As a side remark, we would like to point out that mode of delivery of the source is quite frequently neglected in corpus studies on interpreting, which is a challenge that would need to be addressed in future research.

Looking across modes, comparable corpora help seeing discrepancies between oral and written products of mediation. That this aspect is important was discovered early on by the pioneer of Corpus Interpreting Studies, Miriam Shlesinger, who in her seminal paper “Towards a definition of Interpretese” (2009) used comparable intermodal corpora to analyse the differences between English to Hebrew interpretations and translations carried out in experimental settings. She observed consistent differences in lexical patterns, e.g. lexical variety, which was higher in all written translations. Moreover, simultaneous interpretations used more non-Hebrew borrowings. The general observation regarding the use of register was that interpreters preferred unmarked forms, while more formal lexical choices dominated written translations. In general, she observed that “the features that distinguish spoken Hebrew from written Hebrew are (even) more pronounced in interpreted and translated outputs, respectively” (Shlesinger 2009: 250). Shlesinger’s initial findings to a great extent have been confirmed in more recent intermodal analyses, which also augment her perspective. Interesting intermodal comparisons have also been made with respect to phraseology. Corpora provide means to automatically distinguish between different types of collocations and to further test whether these are used differently across mediation modes. One way to look at it is to use comparable corpora to investigate the way interpreters and translators employ most frequent collocations (usually recognized by high T-score values) and strongly associated collocations (with high mutual information score) that are less frequent. It turns out, for example, that the language of Italian translations is more phraseologically conventional and more idiomatic when compared to the language of Italian interpreters as indicated by higher MI score values of the collocations used in translations than in interpretations. On the other hand, it seems that both interpreters and translators use the most frequent collocations, indicated with higher T-score values, to a similar extent. It is likely in this case that the high cognitive constraints of interpreting affect the choice of fixed and less fixed expressions and that source text interference might play a bigger role in interpreting than in translation (for more information on intermodal investigations into mediated discourse see the work of Adriano Ferraresi).

Different use of syntax and grammar

Apart from the lexical investigations, comparable corpora invite analyses between mediated and non-mediated varieties at the level of grammar and syntax.

First at unimodal level, comparing syntactic and grammatical structures used by interpreters and original oral texts produced by native speakers it is possible to trace patterns favoured by interpreters, especially in specific language pairs and sometimes even to link the complexity of structural patterns to the cognitive load in interpreting. In a deductive, corpus-based approach, such explorations might start with the contrastive analysis and identification of typological differences between the source and target languages, which might be treated as a point of departure for investigations of interference at the syntactic level. A good illustration of such an investigation is the analysis of the interpretation of long and complex attributive modifying structures from Chinese to English. In Chinese, these structures are typically front-loaded, while English is characterized by

back-loaded modifying structures. Despite this difference over 80% of such complex front-loaded structures in Chinese are interpreted into back-loaded structures in English. Wang & Zou (2018: 65) conclude that this means an extra cognitive effort on the part of interpreters linked to the necessity of restructuring while interpreting between so structurally different language pairs. Another case where extra cognitive effort is needed is the interpretation from SVO (subject-verb-object) into SOV (subject-object-verb) languages. One way to investigate this phenomenon with the use of corpora is to look at the so-called “middle field”, i.e. the number of elements between the subject and the verb in the interpreters’ output and non-mediated native speeches. It turns out that interpreters tend to relieve the cognitive effort by shortening the middle field (for more details on this and other corpus-based interpreting studies see the work of Camille Collard).

Second, as in the case of lexis, a lot about syntax and grammar of interpreting and translations can be inferred from intermodal comparisons. For example, interpretations and translations manifest a different use of selected verbal patterns. Shlesinger (2009) found slightly more verb-based constructions in interpretations when compared to translations, consistently different patterns of possessives and definite articles. Gast & Borges (2023) investigated English-German interpretations and translations and their findings confirm in particular the asymmetry in nominal and verbal patterns observed by Shlesinger in the Hebrew-English language pair. Looking at these and other tendencies, e.g. a greater use of adverbs or pronouns they conclude that “interpretations have a hybrid status and can be located somewhere in the middle, between the register of the source text (parliamentary speech) and unplanned spoken discourse”.

A very potent method of corpus exploration involving comparable corpora is multidimensional analysis (for more information on this method consult the work of Douglas Biber), which allows to look at a number of grammatical and stylistic linguistic features at once. It was originally invented to spot differences across registers but has also been used to investigate contact varieties and L2 interpreting and L2 translation (Xu 2021). In very simplistic terms, multidimensional analysis is based on the statistical technique of exploratory factor analysis that allows to group a number of variables otherwise less straightforward to interpret into the, so called, factors/dimensions, which are supposed to reflect the latent relational structure between the features involved. This, on the other hand, can help prepare a multidimensional description of the language used in a collection of texts and compare these text collections to one another with the same ‘yardstick’ (mean dimension score). One of the most prominent sets of features, to a great extent grammatical and syntactic ones, examined in such a way that they are usually grouped in Dimension 1 pertain to involved vs. informational production and usually texts that are delivered orally are placed on one end of this dimension, while written texts on the other. This is also what has been observed by Xu (2021) in an intermodal study, where the key findings concern Dimension 1 and Dimension 2. In terms of Dimension 1, interpreting into L2 was associated with involved and informal features. In this particular study, features loading into Dimension 1 included, among other, present tense, contractions, be as main verb, independent clause coordination wh-pronouns and wh-clauses. Dimension 2, 'On-line information elaboration with stance taking concerns' grouped features which the mediated varieties had in common, contrary to non-mediated varieties. These included, for example amplifiers, demonstratives, that relative clauses on subject positions. According to Xu such shared tendency might also point to common risk management strategies of translators and interpreters. Other dimensions in this analysis provide for a more detailed and comprehensive

linguistic description and the method is increasingly used today in unimodal and intermodal analyses of mediated and constrained texts.

Oral corpora and quality

It looks like some of the linguistic patterns briefly described above can predict to some extent how humans evaluate simultaneous interpreting. This approach was undertaken by Ouyang, Qianxi & Junying (2021), who tried an automatic assessment of interpreting quality. The researchers looked in particular at the word count, lexical diversity, hypernymy of verbs, and frequency of first-person singular pronouns. Their linear regression model accounted for 60% of the variance in the human score: this means that a statistical test, which included the above-mentioned features as fixed variables (predictors) reached the same decision as human raters in about 60% of the data. It was observed that interpretations evaluated highly by human raters tended to use more precise albeit less frequent lexical items and were lexically more diverse.

Parallel corpora

Parallel corpora are a good method to investigate the diverse processes that affect the textual features of the target texts with reference to the source text, such as the tendency to explicitate the information encoded in the source message. Aligned and segmented texts make it easier for the researchers to investigate the correspondence between ST and TT.

There are not many unimodal analyses of parallel interpreting corpora that look at typically linguistic features of the interpreted text. Yet, such a corpus makeup invites the investigations of, for example, interpreting strategies. With adequate corpus annotation, important factors in interpreting, like directionality, position in the source fragment and phraseological richness of the source might be then taken into account in the analysis of interpreting strategies such as, for example, *saucissonage* (also known as a salami technique) or omission. This approach is reported in Dayter (2020) in an analysis of a parallel bidirectional corpus of Russian – English simultaneous interpreting, who finds that in her data the association of the three variables with the type of strategy is significant. If the corpus contains the metadata regarding individual interpreters obtained, for example in the interpreter identification process, idiosyncratic preferences and variation across individual interpreters can be explored. For example, Gumul & Bartłomiejczyk (2022) look at the consistency and frequency of idiosyncratic patterns in search of styles of interpreters reflected in explicitating shifts that affect the cohesion, syntax and texture of the target texts of interpreters' working for the Polish Language Unit at the European Parliament. Additionally, it is worth stressing at this point that corpora involving interpreter identification might encourage a more balanced research design and make it possible to account for idiosyncratic variation in statistical models, which improves the generalizability of results, in particular in the case of such small corpora as interpreting corpora.

Similar explorations of strategies that affect the textual aspects of interpreted output, like cohesion, can be carried out on parallel inter-modal corpora. By consulting the source texts of interpretations and translations it is possible to see to what extent adding and omitting connectives occurs across both mediation modes. Unsurprisingly interpreters tend to omit more most probably due to excessive cognitive load or as a conscious strategy where the connectives are not essential to understand the meaning. Interpreters also add more connectives than translators, which is contrary to expectations,

however, as this might increase the cognitive load (for more information about this and other studies on interpreting using parallel corpora see the work of Bart Defrancq).

Paralinguistic features and quality

Looking at paralinguistic features of orality specific to interpreting, parallel corpora can be used to investigate problem triggers that affect the quality of interpreter's output and source text-determined factors that affect it. This, in turn, can be informative of certain cognitive processes accompanying the interpreting process. By comparing the output of interpreters with the aligned source text in a parallel corpus it is possible to investigate what affects, for example, the fluency of the interpretation. Fluency is frequently operationalised (i.e. measured) with the number of false starts or filled and unfilled pauses or pause length. Careful examinations of these parameters can help identify the most cognitively demanding areas in the source text. This approach has been used for example with respect to interpreting numbers (Kajzer-Wietrzny, Ivaska & Ferraresi 2023) where scholars looked at types of numbers that pose the greatest challenge to interpreters and the context that makes interpreting them more difficult, e.g. the source speech given by a non-native speaker.

The length of silent and filled pauses in the interpreters' output after 10 seconds of the occurrence of less frequent words in the original text, for example, can be measured to investigate the so-called spillover effect, i.e. a decrease in interpreting quality following a problem trigger in the source text. This means that after a more demanding passage the interpreter will have strained their cognitive resources and might provide a less fluent interpretation even if the following fragment of the source speech is not that taxing. The potential problem triggers can also be identified with corpus methods. For example, as shown in the study on spillover effect by Chmiel, Janikowski, Koržinek *et al.* (2023), frequency of words in a reference corpus can be used to determine the potential difficulty of the source text words (in this case there were cognate and non-cognate words).

Parallel corpora also present some potential for the automatic evaluation of interpreting quality, although it has to be stressed that this line of research is still at a very exploratory stage and a fully automated analysis is impossible at this point. However, certain linguistic and para-linguistic information encoded in an interpreting corpus can serve as a useful yardstick to measure the selected aspects of interpreting quality. It looks like the output fluency constitutes a parameter that could be computed and evaluated entirely automatically and that high scores in fluency parameters correspond to higher human quality ratings (Liu 2021: 171).

back to top

Research potential

As we have shown, corpus studies in both translation and interpreting are growing dynamically and they start capturing more and more complexity in increasingly common multifactorial designs extending the scope to novel social and cognitive aspects of translation and interpreting.

For example, in recent years it has been hypothesised that some of the recurring features observed in translated and interpreted texts might not be distinctive of these mediated varieties, but might occur as a result of language contact in other communicative situations, such as non-native production (e.g. learner language, EFL) and other forms of constrained communication (see in

particular Haidee Kotzee's work), due to the presence of similar socio-cognitive mechanisms at work. A more accurate understanding of the mechanisms underlying translation could thus be achieved through more interdisciplinary approaches linking corpus research to neighbouring disciplines like second language acquisition research, bilingualism studies, psycholinguistics and cognitive linguistics.

A relevant example of an area in which the full research potential of CTS and CIS has not yet been sufficiently explored involves translation and interpreting into a foreign language. This issue is linked on the one hand to the cognitively complex L2 processing in mediation, and on the other to the socio-economic aspect of the lack of translators and interpreters into L1 in the case of languages of low or limited diffusion.

Moreover, the vast majority of corpus studies in translation and interpreting is corpus-based with pre-existing operationalisations of common research problems being addressed again and again. Although this might have advantages from the point of view of replication and validation of the observed linguistic patterns on other language pairs/datasets, this also narrows down the range of investigated and re-investigated research questions. A huge potential in this respect lies within corpus-driven approaches which can be helpful in revealing unexpected patterns and lead CTS and CIS to completely new research pathways.

The use of corpora in other subfields of translation and interpreting like audiovisual translation (e.g. Baños, Bruti & Zanotti 2013) or remote interpreting (e.g. Castagnoli and Niemants 2018) is in its infancy. The complexity of annotation associated with corpus work on some forms of language mediation - e.g. sign interpreting involving both text, sound, video alignment and complex indexing of sign systems - makes the compilation of such corpora exceptionally laborious and, for now, seriously affects their size. Research on sign language interpreting, however, is a particularly dynamically developing field and the number of sign language corpus resources is constantly growing. One can hope that with the fast development of artificial intelligence, the compilation and annotation of these corpora will require less human effort.

The compilation of some corpora is more laborious than others. The more communication modes are involved, the more complex the annotation scheme. Following this general rule, the compilation of written translation corpora is usually less laborious than the compilation of interpreting, audiovisual or sign language corpora, in particular. Some data is also more difficult to obtain due to copyright or anonymisation issues (anonymising sign language recordings or recorded simultaneous interpretations might be more challenging than anonymising texts) among other things.

Some of the above difficulties might be tackled by pulling together the expertise and resources of different research teams from several institutions, which for now is not a common practice among scholars engaged in corpus studies on translation and interpreting. There are notable exceptions, involving data sharing and combining corpus resources from different projects like in the case of EPIC UDS corpus, where researchers reached out to colleagues to expand the set of interpreting transcripts at their disposal. An intermodal corpus, EPTIC, is another example of the ongoing collaboration of interpreting and translation scholars from different universities across Europe and the researchers are making the outcomes of this collaboration available online. Yet another showcase of scholarly cooperation in corpus creation is the MUST project, which has so far resulted in the compilation of a 2 million-token corpus of student translations (available, for now, only to the

MUST community) richly annotated with metadata and student translation errors following a dedicated error taxonomy. Such initiatives prove that inter-institutional collaboration is a successful way to overcome data scarcity, which is an infallible way to make corpus studies in translation and interpreting as well as their outcomes more robust.

back to top

Bibliography

Baker, Mona. 1993. “Corpus Linguistics and Translation Studies - Implications and Applications”. @ Baker Mona, Gill Francis & Elena Tognini-Bonelli (eds.) 1993. *Text and Technology - In Honour of John Sinclair*, 233-250. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/z.64.15bak [+info]

Baños, Rocío; Silvia Bruti & Serenella Zanotti. 2013. “Corpus linguistics and audiovisual translation: in search of an integrated approach”. @ *Perspectives* 21/4, 483-490. DOI: https://doi.org/10.1080/0907676X.2013.831926 [+info]

Bernardini, Silvia & Adriano Ferraresi. 2011. “Practice, description and theory come together – normalization or interference in Italian technical translation?”. @ *Meta* 56/2, 226-246. DOI: https://doi.org/10.7202/1006174ar [+info] [quod vide]

Cappelle, Bert & Rudy Loock. 2017. “Typological differences shining through. The case of phrasal verbs in translated English”. @ De Sutter, Gert, Marie-Aude Lefer & Isabelle Delaere (eds.) 2017. *Empirical translation studies. New methodological and theoretical traditions,* 235-264. Berlin: De Gruyter Mouton. [+info]

Carl, Michael & Moritz Schaeffer. 2017. “Measuring translation literality”. @ Jakobsen, Arnt Lykke & Bartolomé Mesa-Lao (eds.) 2017. *Transla-tion in Transition: Between cognition, computing and technology*, 81-106. Amsterdam: John Benjamins. [+info]

Castagnoli, Sara. 2022. *Learner translation corpora. Exploring regularities and variation in student translation*. Roma: Aracne. ISBN: 9791221802207. [+info]

Castagnoli, Sara & Natacha S. A. Niemants. 2018. “Corpora worth creating: A pilot study on telephone interpreting”. @ *inTRAlinea* 20 - *Special Issue: New Findings in Corpus-based Interpreting Studies* [s.p.]. [+info] [quod vide]

Chesterman, Andrew. 2004. “Beyond the particular”. @ Mauranen, Anna & Pekka Kujamäki (eds.) 2004. *Translation universals – Do they exist?,* 33-49. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/btl.48.04che [+info]

Chmiel, Agnieszka; Przemysław Janikowski; Danijel Koržinek; Agnieszka Lijewska; Marta Kajzer-Wietrzny; Dariusz Jakubowski & Koen Plevoets. 2023. “Lexical frequency modulates current cognitive load, but triggers no spillover effect in interpreting”. @ *Perspectives* 1-19. DOI: https://doi.org/10.1080/0907676X.2023.2218553. Online preprint. [+info]

Czulo, Oliver; Silvia Hansen-Schirra & Jean Nitzke. 2017. “Contrasting terminological variation in post-editing and human translation of texts from the technical and medical domain”. @ De Sutter, Gert; Marie-Aude Lefer & Isabelle Delaere (eds.) 2017. *Empirical translation studies. New methodological and theoretical traditions,* 183-206. Berlin: De Gruyter Mouton. [+info]

Dayter, Daria. 2020. “Strategies in a corpus of simultaneous interpreting. Effects of directionality, phraseological richness, and position in speech event”. @ *Meta* 65/3, 594-617. DOI: https://doi.org/10.7202/1077405ar [+info] [quod vide]

* De Sutter, Gert; Marie-Aude Lefer & Isabelle Delaere (eds.). 2017. *Empirical translation studies. New methodological and theoretical traditions*. Berlin: De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110459586 [+info]

De Sutter, Gert & Marie-Aude Lefer. 2020. “On the need for a new research agenda for corpus-based translation studies: a multi-methodological, multifactorial and interdisciplinary approach”. @ *Perspectives* 28/1, 1-23, DOI: 10.1080/0907676X.2019.1611891 [+info]

Delaere, Isabelle; Gert De Sutter & Koen Plevoets. 2012. “Is translated language more standardized than non-translated language? Using profile-based correspondence analysis for measuring linguistic distances between language varieties”. @ *Target* 24/2, 203-224. DOI: https://doi.org/10.1075/target.24.2.01del [+info]

Dragsted, Barbara. 2012. “Indicators of difficulty in translation – Correlating product and process data”. @ *Across languages and cultures* 13/1, 81-98. DOI: https://doi.org/10.1556/Acr.13.2012.1.5 [+info]

Gast, Volker & Robert Borges. 2023. “Nouns, Verbs and Other Parts of Speech in Translation and Interpreting: Evidence from English Speeches Made in the European Parliament and Their German Translations and Interpretations”. @ *Languages* 8/1, Article 1. DOI: https://doi.org/10.3390/languages8010039 [+info] [quod vide]

Gile, Daniel. 2004. “Translation Research versus Interpreting Research: Kinship, Differences and Prospects for Partnership”. @ Schäffner, Christina (ed.) 2004. *Translation Research and Interpreting Research,* 10-34. Bristol: Multilingual Matters. DOI: https://doi.org/10.21832/9781853597350-003 [+info]

Gumul, Ewa & Magdalena Bartłomiejczyk. 2022. “Interpreters’ explicitating styles: A corpus study of material from the European Parliament”. @ *Interpreting* 24/2, 163-191. DOI: https://doi.org/10.1075/intp.00081.gum [+info]

Halverson, Sandra L. 2017. “Gravitational pull in translation: Testing a revised model”. @ De Sutter, Gert; Marie-Aude Lefer & Isabelle Delaere (eds.) 2017. *Empirical translation studies. New methodological and theoretical traditions,* 9-46. Berlin: De Gruyter Mouton. [+info]

Kajzer-Wietrzny, Marta. 2015. “Simplification in interpreting and translation”. @ *Across Languages and Cultures* 16/2, 233-255. DOI: https://doi.org/10.1556/084.2015.16.2.5 [+info]

Kajzer-Wietrzny, Marta; Ilmari Ivaska & Adriano Ferraresi. 2023. “Fluency in rendering numbers in simultaneous interpreting”. @ *Interpreting*. DOI: https://doi.org/10.1075/intp.00092.kaj Online preprint. [+info]

Kunilovskaya, Maria; Tatyana Ilyushchenya; Natalia Morgoun & Ruslan Mitkov. 2022. “Source language difficulties in learner translation: Evidence from an error-annotated corpus”. @ *Target* 35/1, 34-62. DOI: https://doi.org/10.1075/target.20189.kun [+info]

Lapshinova-Koltunski, Ekaterina. 2022. “Detecting normalisation and shining-through in novice and professional translations”. @ Granger, Sylviane & Marie-Aude Lefer (eds.) 2022. *Extending the scope of corpus-based translation studies,* 182-206. London: Bloomsbury. [+info]

Laviosa, Sara. 1998. “Core patterns of lexical use in a comparable corpus of English narrative prose”. @ *Meta* 43/4, 557-570. DOI: https://doi.org/10.7202/003425ar [+info] [quod vide]

Li, Defeng; Chunling Zhang & Kanglong Liu. 2011. “Translation style and ideology: A corpus-assisted analysis of two English translations of Hongloumeng”. @ *Literary and Linguistic Computing* 26/2, 153-166. DOI: https://doi.org/10.1093/llc/fqr001 [+info] [quod vide]

Liu, Yanmeng. 2021. “Exploring a corpus-based approach to assessing interpreting quality”. @ Chen, Jing & Chao Han (eds.) 2021. *Testing and assessment of interpreting: Recent developments in China*, 159-178. Singapore: Springer. [+info]

Mikhailov, Mikhail & Miia Villikka. 2001. “Is there such a thing as a translator’s style?”. @ [n.n.] 2001. *Proceedings of the Corpus Linguistics conference 2001*. UCREL Technical Papers 13, Special issue, 378-385. Lancaster: Lancaster University. [+info]

Oakes, Michael P. & Meng Ji (eds.) 2012. *Quantitative Methods in Corpus-Based Translation Studies*. Amsterdam: John Benjamins. [+info]

* Olohan, Maeve. 2005. *Introducing corpora in translation studies.* London: Routledge. [+info]

Ouyang, Lingwei; Lv Qianxi & Liang Junying. 2021. “Coh-Metrix model-based automatic assessment of interpreting quality”. @ Chen, Jing & Chao Han (eds.) 2021. *Testing and assessment of interpreting: Recent developments in China,* 179-200. Singapore: Springer. DOI: https://doi.org/10.1007/978-981-15-8554-8_9 [+info]

Popovic, Maja; Ekaterina Lapshinova-Koltunski & Maarit Koponen. 2023. “Computational analysis of different translations: by professionals, students and machines”. @ [n.n.] 2023. *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, 365-374. Tampere, Finland. [+info] [quod vide]

Russo, Mariachiara; Claudio Bendazzoli & Annalisa Sandrelli. 2006. “Looking for lexical patterns in a trilingual corpus of source and interpreted speeches: Extended analysis of EPIC (European Parliament Interpreting Corpus)”. @ *FORUM. Revue Internationale d’interprétation et de Traduction/International Journal of Interpretation and Translation* 4/1, 221-254. [+info]

Saldanha, Gabriela. 2011. “Translator style: Methodological considerations”. @ *The translator* 17/1, 25-50. DOI: https://doi.org/10.1080/13556509.2011.10799478 [+info]

Shlesinger, Miriam. 2009. “Towards a definition of Interpretese: An intermodal, corpus-based study”. @ Hansen, Gyde; Andrew Chesterman & Heidrun Gerzymisch-Arbogast (eds.) 2009. *Efforts and Models in Interpreting and Translation Research: A tribute to Daniel Gile*, 237-253. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/btl.80.18shl [+info]

Tirkkonen-Condit, Sonja. 2004. “Unique items — over- or under-represented in translated language?”. @ Mauranen, Anna & Pekka Kujamäki (eds.) 2004. *Translation universals – Do they*

exist?, 177-184. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/btl.48.14tir [+info]

Toury, Gideon. 1981. “Translated Literature: System, Norm, Performance. Toward a TT-Oriented Approach to Literary Translation”. @ *Poetics Today* 2/4, 9-27. [+info]

Toury, Gideon. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins. [+info]

Wang, Binhua & Bing Zou. 2018. “Exploring Language Specificity as a Variable in Chinese-English Interpreting. A Corpus-Based Investigation”. @ Russo, Maria Chiara; Claudio Bendazzoli & Bart Defrancq (eds.) 2018. *Making Way in Corpus-based Interpreting Studies*, 65-82. Singapore: Springer. DOI: https://doi.org/10.1007/978-981-10-6199-8_4 [+info]

Winters, Marion. 2009. “Modal particles explained: How modal particles creep into translations and reveal translators’ styles”. @ *Target* 21/1, 74-97. DOI: https://doi.org/10.1075/target.21.1.04win [+info]

Xu, Cui. 2021. *Identification of L2 interpretese: A corpus-based, intermodal, and multidimensional analysis*. Hong Kong: Hong Kong Polytechnic University. Doctoral thesis. [+info] [quod vide]

Credits

Sara Castagnoli

Associate professor in English language and translation at the Department of Education, Cultural heritage and Tourism of the University of Macerata, Italy. Her research activity spans several areas of corpus-based translation studies, with a focus on learner translation (*Learner Translation Corpora – Exploring regularities and variation in student translation*, Aracne 2022). She has taken part to several national and international research projects focusing on the construction and use of corpora for professional, teaching and research purposes, including lexicographic and terminological applications.

Marta Kajzer-Wietrzny

Assistant professor in the Department of Translation Studies at the Faculty of English, Adam Mickiewicz University in Poznań. Following her PhD dissertation on *Interpreting universals and interpreting style* (2012) she continues with empirical investigations of mediated language e.g. within the TRINFO project carried out in part during an over year-long research stay at the University of Bologna or as a team member of the PINC project run at Adam Mickiewicz University. At times she attempts to combine corpus methods with process research such as key-logging and eye-tracking in investigations on both interpreting and translation.

Asociación Ibérica de Estudios de Traducción e Interpretación (AIETI)