

Metaphorical keyness in specialised corpora

Gill Philip

University of Bologna, Italy

What is more important in text: the topical content, or the manner in which topical content is presented? While statistically-generated key words tell us about a text's content, the *inter-relation between these words and the message* of the text can be difficult to ascertain. One method of doing so is to observe the inter-relation of key words with evaluative language: in this case, metaphor. Metaphors are notoriously difficult to locate in corpora, but this paper sets out a method for their semi-automatic identification, and demonstrates how their interaction with keywords is both systematic and pervasive. Studying the interaction of key words and metaphors brings to light attitudes which lurk beneath the surface of text.

1. Introduction

The function of a key is to open locks; and locks are used to ensure the safe-keeping of those things we consider valuable, precious, or important. The identification of key words in a text provides the reader with the opportunity to gain direct, unfettered access to its content, circumventing the need to pick the lock (by reading the full text in detail, then analysing its contents) or peer through the keyhole (by making generalisations based on its most striking features). The identification of key words in a specialised corpus reveals textual information on a larger scale, highlighting the matters which the overall discourse is concerned with. Yet while key words tell us *what* is important, they do not tell us *why*.

It is not necessary for a word to appear at the top of the word-frequency list for it to qualify as key, just as not all high-frequency vocabulary necessarily has key word status (see Appendix 1). However, it must recur: a discourse cannot hinge upon hapax legomena. For this reason, the calculation of key words is dependent on frequency measures and repetition, yet these matters are not entirely unproblematic. In particular, a language with very few inflected forms has more recurrent forms than a fully inflected one, which in turn has fewer forms

than agglutinative or infixing languages. While each word form attracts its own distinctive patterning, the dispersion of closely-related meanings over variant forms of a lemma may affect frequency measures and statistical calculations¹. A further matter, and one which is the central concern of this paper, is that there are some kinds of repetition which do not involve replication or reiteration of the same word forms. This is true of semantic relations in particular, and though semantic annotation tools can aid the recognition of related senses (Rayson 2005), they are limited both in terms of availability for languages other than English, and in the types of relation that they can uncover². The grouping of semantically-related words into lexical sets implies recurrence, but not of a sort which can be measured in a straightforward way. Lexical sets can form around a key word, thus acquiring the status of *key-by-association*, and these add richness to the lexis regarding a text or discourse topic. Many other lexical sets may also occur, but their presence is less easily noticed. These sets lurk in low-frequency vocabulary; they do not seem to be relevant to the aboutness of the discourse, indeed they may seem out of place. These are the groupings that suggest metaphorical activity.

At first glance, the relevance of metaphor to notions of keyness might appear minimal, as keyness is concerned with subject matter, and, as a general rule, metaphors are not informative nor are they central to the transmission of content³. What metaphors are used for, however, is the expression of “affect and attitude along with ideational content” (Cameron & Deignan 2006:676), meaning that metaphors play an important evaluative role in a text or discourse. This is as true of metaphors which are deliberately used for their rhetorical function as it is for language which is less actively metaphorical and not used with deliberate rhetorical intent – conventional and delexical expressions. Both these types of metaphor interact with high-frequency lexis and key words, and they do so in different,

1. In this study, the language of the data is Italian; thus nouns, verbs, and adjectives are inflected. There are also variant forms for articles and determiners, and fused forms for certain pronominal constructions and for article + preposition.

2. If the lexis is related taxonomically, semantic annotation poses few problems; however conventional notions of semantic relatedness are less effective in the treatment of metaphor and other figurative language, where relations tend to be based on attributes and characteristics (Glucksberg & Keysar 1993).

3. Exceptions to this rule include dead metaphors and metaphorically-motivated terminology whose meaning is considered “basic” (or “literal”) in the language as a whole, or in a given discourse; and explicative metaphors and analogies whose function is to shed light on new information by comparison with something more familiar.

but complementary ways; however it is the latter that form the main focus of this study. Conventional, delexical metaphors recur almost invisibly over many texts. The reader's attention is rarely drawn to them and yet, at an almost subliminal level, they contribute to the meaning of the high-frequency lexis and key words. If a better understanding of the *whys* of key words is sought, the analysis of recurrent metaphors can be fruitful. This paper discusses the location of metaphors in a corpus of specialised language, and the identification of "metaphor themes" (Black 1993; see 2.1) which run through the discourse. The more pervasive of these – the *key metaphors* – are then studied to determine their evaluative power over the statistically-generated key words in the same data set.

2. Metaphor and textual meaning

2.1 Terminology note

The study of metaphor assumes the knowledge of some core terminology. This paper looks at single instances of metaphors – *linguistic metaphors* (after Steen 1994) in the corpus data, and groupings of these linguistic metaphors into semantically related areas, known in the cognitive linguistics literature as *conceptual metaphors* (Lakoff & Johnson 1980); in this paper I adopt the theory-neutral term *metaphor theme* (after Black 1993).

In all metaphors, one entity, idea, or action is described, defined or explained through another. The relationship between these parts is described for linguistic metaphors in terms of *vehicle*, *target*, and *ground* (after Richards 1936), while for metaphor themes it is described in terms of *source* and *target* domains (after Lakoff & Johnson 1980). The lexical items (in the case of linguistic metaphors) or semantic domains (for metaphor themes) must be distinct, and metaphoricity is created because of the incongruity of their use together. In the linguistic metaphor, *il mondo è [...] la nostra ancora di salvezza*⁴ [the world is our safety anchor], there is incongruity between *mondo* [world] (the metaphor target), and *ancora di salvezza* [safety anchor] (the metaphor vehicle), and the ways in which the world can be conceptualised in terms of anchors forms the motivation (ground) for the metaphor. In a conceptual metaphor or metaphor theme, the metaphorical activity is no longer tied to specific words, but rather occurs in the abstract: the target domain (realised through a variety of possible words) is expressed in terms of the source

4. Unless otherwise stated, all examples are from the ComInt corpus (see 3.1).

domain (again realised through a variety of possible words)⁵. Metaphor themes are identified in text by grouping together apparently related linguistic metaphors, which may or may not feature the same vocabulary. Figure 1a shows the **ECONOMIC PRODUCTIVITY IS A HEALTHY BODY** metaphor⁶, illustrated by *paziente convalescente ma robusto* (convalescent but strong patient), *diagnosi* (diagnosis) *sano* (healthy) and *stato di salute* (condition of health), while **INTERNATIONAL TRADE IS WAR** is illustrated in Figure 1b through the lexical items *agguerrita* (ready for war, fierce) *conquista* (conquer), *vincenti* (winning) and *battaglia* (battle).

2.2 Metaphors in text

While literary metaphors are highly visible in text because of their novelty, the same cannot be said of non-literary metaphors. Most metaphors are in fact naturalised lexical items, forming part of the conventional vocabulary and phraseology of the language, and their metaphorical nature tends to pass unnoticed, both in production and in reception. In common with all lexical items, conventional metaphors attract distinct collocational patternings which delineate their meanings (Deignan 1999), and the extended unit of meaning (Sinclair 1996) of a metaphor is no different to any other in that it has semantic preferences and, crucially, semantic prosodies (Philip in press a). These are often invisible but, following a pragmatic approach to metaphor (after Charteris-Black 2004), this invisibility appears to be central to the rhetorical and persuasive use of metaphor in non-literary texts. While Louw (2000) is justified in arguing that the use of conventional metaphor is not indicative of active metaphorical conceptualising, the fact remains that conventional, delexical forms transmit evaluative meaning, irrespective of whether they are consciously produced (or understood) as figures of speech. The repeated use of particular metaphors or metaphor themes, especially when these are used with reference to a limited range of subjects, conveys a sense of the underlying attitudes regarding the subject in hand. While the metaphors are indeed virtually invisible, their presence can be perceived subliminally.

5. In a comparable account of metaphor, “above the actual metaphor as a speech act, in our linguistic competence there is an image field as a visual structure.” (Weinrich 1967:283; this translation cited in Jäkel 1999: 18). In this account, the source domain is labelled the image donor field, and the target domain is the image recipient field (Weinrich 1958:284). However, as the degree to which metaphors are visualised remains subject to individual variation, metaphor theme is considered the most satisfactory term for use in the present study.

6. Metaphor themes are conventionally notated in small capitals. In the examples, the target domain is underlined, and the source is in bold face.

-
- (i) Importanti centri studi hanno fotografato l'Italia economica come un paziente convalescente ma robusto.
'Important study centres have pictured economic Italy as a convalescent, but strong, patient.'
- (ii) La diagnosi che mi sento di condividere è che il sistema produttivo italiano è un sistema sano...
'The diagnosis that I feel able to share with you is that the Italian productive system is a healthy one.'
- (iii) Il buono stato di salute del cinema italiano ha trovato risposte non solo nei paesi da sempre interessati...
'The good health of Italian cinema has been noted not only in those countries which have always expressed an interest...'
-

Figure 1a. ECONOMIC PRODUCTIVITY IS A HEALTHY BODY

-
- (i) Il che significa che, in presenza di una concorrenza agguerrita, noi siamo soccombenti per motivi di costi...
'Which means that, in the presence of fierce competition, we are the losing party because of prices.'
- (ii) Il Made in Italy conquista nuovi mercati, Russia e Cina, ma preoccupa flessione in Usa e Giappone.
'The Made in Italy [brand] is conquering new markets, Russia and China, but is suffering a downturn in the USA and Japan.'
- (iii) Sui mercati mondiali siamo vincenti solo se riusciamo a trasformare la nostra migliore tradizione artigiana...
'We can be winners in international markets only if we manage to transform the best of our tradition of workmanship...'
- (iv) ...nella battaglia per affermare l'Italia sui mercati internazionali.
'...in the battle to assert Italy's place in international markets.'
-

Figure 1b. (INTERNATIONAL) TRADE IS WAR

Non-literary metaphors in text rarely occur in splendid isolation (Example 1), because they are a core part of the vocabulary of any language. Even in combination, when metaphors come in *clusters* (Cameron & Stelma 2004), their figurative nature tends not to be registered consciously. This is because the lexis is being used in formulaic ways which give the reader no reason to stop and evaluate whether any hidden or extended meanings are inferred (Example 2). The matter is somewhat different when the clustering involves lexis drawn from a single lexical set/conceptual domain which is clearly not congruous with that of the discourse topic (as in Example 3). In these instances the lexis, which is normally interpreted with its delexical or textual meaning, undergoes double processing: the words'

proximity and incongruity with the discourse trigger their reinterpretation as simultaneously literal and figurative, with the result that their metaphoricity is accentuated (Philip in press a). However these clusters, or conceits, are relatively rare (only a dozen instances located in the corpus studied), and their use seems deliberate in all instances.

- (1) *È un processo delicato ma importante, tuttora in fase di gestazione...*
'It is a delicate but important process, still in gestation...'
- (2) *...e mi auguro che le scelte che faremo non siano soffocate da miopi atteggiamenti politici di corto respiro.*
'...and I hope that the choices we make will not be suffocated by short-sighted, short-lived politicising.'
- (3) *Il mondo non è il posto dove rischiamo di naufragare, ma la nostra ancora di salvezza contro i rischi di impaludamento che corriamo se restiamo nei nostri piccoli mercati locali.*
'The world not is the place where we risk getting shipwrecked, but our safety anchor against the risk we run of getting stuck in a marsh if we remain in our small local markets.'

Less deliberate, less noticeable, but perhaps more influential in the long run are metaphor themes. Conceits exert considerable influence within single texts, yet have little or no lasting influence over the discourse as a whole. Instances of metaphor themes, on the other hand, are relatively inconsequential in themselves but rise in importance with each subsequent reappearance in the discourse. The recurrence of particular themes over many texts produced or delivered by the same individual can reveal information concerning that individual's stance on the subject matter. Aspects of the broader socio-political climate in which the texts were produced may also surface, especially if the texts belong to a fixed time-frame. If the author or deliverer of the texts is a public figure, such as a politician, the relevance of stance and evaluation is magnified: thanks to the media's habit of reporting politicians' words verbatim, the phraseology used by one becomes the phraseology used by many, as the underlying implications of the politician's linguistic choices resonate beyond the original text into the language at large. Being conventional and unremarkable, these metaphors operate in silence, yet they help to shape the opinions of millions.

3. Metaphors and corpora

If metaphors have the potential to persuade as well as to evaluate, their presence merits attention. However, due to the fact that instances of metaphor themes are spread over many texts, their identification is problematic as well as time-consuming and laborious. Even bringing the texts together as a corpus of specialised language falls short of successfully addressing the problem, because query software is designed for word searches, not meaning searches.

Metaphor scholars working with corpora (Charteris-Black 2004, 2005; Partington 2003; Koller & Semino 2009) have tended to overcome this problem by engaging in partial or total manual searching of the corpus texts, then using computer search tools at the second stage of analysis. The first stage involves close, word-by-word reading of the text, assessing whether each lexical item is used metaphorically or otherwise⁷, in accordance with the established metaphor identification procedure (Pragglejaz Group 2007). The second stage involves calling up concordances of all instances of the identified words and expressions (following the type of analysis proposed by Deignan 1999, 2005: Chapter 4), identification of metaphorical senses, then analysis of these with or without reference back to the extended context or full text in which the metaphors occur. The methods used in the present study take a different approach because the aim is not to locate all metaphors present in the corpus but to identify potential metaphorical themes. The data and methods are set out below.

3.1 Corpus composition

The corpus described in this study is one of six corpora compiled for a study of metaphor use by women Ministers in Italy (Philip 2009). The data was compiled from the speeches, communiqués and press releases produced by Emma Bonino, in her dual role as Minister for International Trade and Commerce, and Minister for European Policy (represented by the ComInt and PolEur corpora respectively). The corpus data covers the first year (June 2006 – May 2007) of the then centre-left government and the data was freely available for download from the Ministerial homepage. The corpus contains approximately 140,000 words of running text divided into four subcorpora, the details of which appear in Table 1.

7. This is determined as follows: "If the lexical unit has a more basic current-contemporary meaning in other contexts than the given context, decide whether the contextual meaning contrasts with the basic meaning but can be understood in comparison with it." (Pragglejaz Group 2007: 3).

Table 1. Corpus composition

	overall	PolEur		ComInt	
		sp.*	p.r.	sp.	p.r.
no. of texts	203	13	59	55	76
mean text length	688	2180	221	1291	358
tokens in text	139,605	28,344	13,057	71,024	27,180
tokens in word list*	136,788	27,832	12,704	70,155	26,097
types	11,865	4,727	2,704	8,948	3,795

* sp. – speeches and speech notes; p.r. – press releases

3.2 Locating metaphorical lexis

The methodology set out here was devised in order to identify metaphors in the corpus data without reading through and annotating the texts beforehand. Although the manual analysis of a small (<150,000 words) corpus is feasible, it is time-consuming and prey to inconsistencies, omission and misclassifications due to human error. A need therefore arises for a semi-automatic procedure for identifying metaphors, which eliminates the need to analyse a core of the data beforehand, and leads directly into concordance analysis. The procedure described was carried out using WordSmith Tools version 4 (Scott 2004), but can be carried out with any concordance package which generates word frequency and key word lists.

The most important feature of metaphor, as noted in 2, is that it requires the presence of contextually incongruous lexis, and this central fact about metaphor informs the method used in this study. If metaphorical lexis is not central to the topic of the discourse, it will not occur near the top of a word-frequency list, nor will it feature amongst the statistically-generated key words, but will stand in contrast to it. With this knowledge as a starting point, a word frequency list was created for the ComInt corpus, and a key word list was generated using the word frequencies of the entire political corpus data (430,000 words) as the control. The key words were thus identified with reference to the political discourse of the time-frame in which the data was collected, not to the language overall. This made it possible to identify the topic-related lexical areas of the specialised corpus, while eliminating lexis common to political language in general, such as 'Minister', 'country', 'government'; these appear high up on the frequency list, but are not key (see Appendix 1). The key words were lemmatised where appropriate, then grouped into semantic sets in order to establish the criterion of incongruity, necessary for metaphor identification. The groups identified were: markets and

sectors; business and industry; the economy; import and export, foreign countries and internationalisation; legislation; and development (see Appendix 2).

Some metaphor themes are known to be typical of certain discourses. In business and economics, these include orientation metaphors (ups and downs), metaphors of growth, relationships, and water, amongst others. In the key word list, two such metaphors occurred, namely *crescita* [growth] and *flussi* [flow]. Both also occurred as high-frequency lexis, suggesting that they are not metaphors but rather metaphorically-motivated terminology. As metaphors are created by contrasting basic (literal, terminological, salient) meanings with an incongruous domain, it is highly unlikely that a word form will be used both figuratively and literally within the same text or discourse. No near-synonyms were found for *crescita* and *flussi*, confirming the hypothesis that these metaphors have terminological status. The absence of related forms makes it impossible to ascertain whether terminological status is limited to these lemmas or whether it also extends to the semantic domain to which each belongs. Both *crescita* and *flussi* were excluded from consideration as possible metaphor themes in this corpus.

While high-frequency content words sum up the aboutness of the data, it was hypothesised that low-frequency content words (LFCWs) would provide a source of metaphorically-used lexis. There are two obvious problems with this observation. The first is that LFCWs make up a very large proportion of any corpus – in ComInt, hapax legomena alone account for over 40% of all word forms (see Table 2). The second problem is that not all low-frequency lexis is metaphorical.

Table 2. Details of lowest-frequency word forms (bottom 15% of tokens)

no. *	rank range	types		tokens	
		total	%	total	%
1	11,865–6805	5060	42.65	5060	3.70
2	6804–3679	3125	26.35	6250	4.55
3	3678–2854	753	6.35	2259	1.65
4	2853–2340	513	4.30	2052	1.50
5	2339–1998	341	2.90	1705	1.25
6	1997–1716	281	2.40	1686	1.25
7	1715–1506	209	1.77	1463	1.05
	overall	10,202	86.66	20,475	14.95

* number of occurrences of each type

The analysis of the ComInt data started by (manually) lemmatising the top of the word frequency list in order to eliminate low-frequency inflected forms which, when grouped together, constituted medium- to high-frequency lemmas. This

process concerned the top 500 types and their co-inflected forms, this slice of the data being the limit beyond which no key-words were located. The remaining lexis was sorted into very broadly-defined semantic groups, starting at the bottom of the list with the hapax legomena then proceeding upwards, with an initial cut-off point established at 3 occurrences (this is equal to approximately 75% of the running words in the corpus). This cut-off point allowed for the identification of semantic groups which were sufficiently distinct from those represented by the key words so as to be classed as potential metaphorical source domains (a semantic group was defined as such if it contained five or more lemmas; fewer than five proved to be insufficient grounds for classification). Lemmas which could not be grouped with others, such as those listed in Table 3, were considered as candidates for linguistic metaphors, but not metaphor themes.

Table 3. Selection of ungrouped metaphor candidates

baricentro [centre of gravity]	cabina [cabin]
catene [chains]	clima [climate]
fetta [slice]	guai [trouble]
labbra [lips]	lupo [wolf]
naufragare [shipwreck]	ombelico [navel]
orizzonti [horizons]	riva [(river) bank]
scale [stairs]	seno [breast]

The higher the frequency of a metaphor candidate, the more consistency there was in its patterning in the corpus overall. The crystallisation of patternings around the node word was observed at >5, and at 7 occurrences was already marked in some cases. Strong collocational and phraseological preferences affect the polysemous potential of a word, limiting the likelihood that it will be used both literally and figuratively in the same discourse. As different meanings imply different patterns, the emergence of dominant patternings in a text or discourse makes it less likely that other patterns – and hence, other senses – will occur. This hypothesis is confirmed by the data in the corpus, where evidence is found of the crystallisation of collocational patternings, visible with upwards of ten concordance lines for the same word form, and occasionally with even fewer.

Figure 2 shows the concordance lines for *penetrazione* (penetration), where it can be seen that there is a preferred collocate *commerciale* (commercial), and a less striking but nonetheless visible preference for *penetrazione* to concern markets and sectors (*mercati, settori*), abroad rather than at home (*esteri, internazionali*). Metaphor candidates occurring in the middle-frequency bands (below the key word threshold, and, in this corpus, above ten occurrences), demonstrate greater cotextual stability than the lower-frequency candidates, and as a result begin to

a punto un piano strategico di	penetrazione	commerciale dal 2008 al 2010 , indivi
delle azioni di sostegno alla	penetrazione	commerciale del sistema Italia. Siamo
capacità di esportazione e di	penetrazione	commerciale dei nostri imprenditori e
no strumento strategico per la	penetrazione	commerciale delle nostre imprese. I m
utti , per la maggior parte di	penetrazione	commerciale finanziati attraverso la
re operazioni più complesse di	penetrazione	commerciale. Mi auguro che questi nuo
ia attività di promozione e di	penetrazione	commerciale. Per l ' anno 2007 il Min
so di internazionalizzazione e	penetrazione	commerciale. Il Ministero del Commer
e sue possibilità di ulteriore	penetrazione	commerciale su mercati maturi ma anch
ed iniziativa che rafforzi la	penetrazione	delle imprese editoriali italiane nei
lo di forte protagonismo nella	penetrazione	dei mercati esteri. Nella situazione
la meccanica strumentale , una	penetrazione	nel settore dei servizi e in quelli a
nazionali , di accompagnare la	penetrazione	sui mercati internazionali con adegua

Figure 2. Crystallisation of patterning around 'penetrazione': All occurrences

consolidate themselves as domain-specific vocabulary or indeed terminology (Philip in press b), but this aspect is beyond the scope of the present study⁸.

While it is true that metaphorical activity is determined by phraseology, and that word counts fail to distinguish between different word senses, the method described above is designed to aid the identification of metaphor candidates, not as solid proof of metaphorical activity. The words identified as potentially metaphorical were concordanced, and on the basis of the cotextual evidence, literal uses were discarded from the data set.

4. Metaphor themes and key metaphor themes

The analysis described in 3.2 yielded a large set of word forms that constitute a broad category of war and violence, plus smaller groups representing hunting, risk, submission and suffering, health, birth, death, and emotion. These areas constitute the source domains identified, but they do not become metaphor themes until the target domains have been ascertained. It is not enough to say that BUSINESS IS WAR, or RISK, or A LIVING ORGANISM: these themes are too general and of limited value in text analysis. The metaphorical targets are identified by concordancing each of the word forms separately, which makes it possible to arrive at more detailed metaphor themes which specify, for instance, that ECONOMIC PRODUCTIVITY IS A HEALTHY BODY (c.f. Figure 1a) or that INTERNATIONAL TRADE IS WAR.

8. For detailed treatment and evaluation of the methodology used in this paper, see Philip (in press b).

War metaphors are the most frequent in this corpus, and they are the one theme that seems to be common to all political discourse (Philip 2009). Identifying the metaphor targets through corpus analysis makes it possible to penetrate the mass of apparently common metaphors and home in on domain-specific themes and their uses. By looking in detail at the use of the source domain lexis, it was established that the lexical set comprising battles, fights and skirmishes was specifically linked to foreign trade, with particular reference to the main emerging economies, China and India. However, the different aspects of this state of affairs are expressed by specific word choices, meaning that the source domain terms are not interchangeable: the *battaglia* [battle] is for Italy to maintain its competitive advantage in world markets; the luxury *Made in Italy* brand is engaged in a *lotta* [fight, struggle] against the influx of imitation and counterfeit goods. Consistent with the reference to foreign trade, this metaphor also includes invasion of foreign territories, and once again the source-domain words are not interchangeable: *invadere* [invade] and *penetrare* [penetrate] are effectively synonymous, but while Italy's expansion into foreign markets (especially China and India) is expressed as *penetrazione* [penetration], the expansion of those same nations into Europe (and Italy in particular) is described as *invasione* [invasion] against which the reputation of *Made in Italy* products must be defended (Philip 2009: 105).

It is not the ubiquity of the war metaphor that makes it a key metaphor. Key metaphors are key not because they are the most frequently used within a text or discourse, but because they interact in significant ways with the key words. An apparently pervasive source domain may be found to disintegrate into several relatively infrequent metaphor themes, reducing its status. A less frequent source domain which is mapped onto just one target may, on the other hand, prove to play a more significant evaluative role in the discourse. The case of the war metaphors in this corpus is different again: as a source domain, war is by far the richest, comprising four times the number of word forms (112) assigned to the next-most frequent grouping, life and death (34). The subdivision of this domain into specific metaphor themes yields INTERNATIONAL TRADE IS WAR, which can be subdivided into TRADE IS AGGRESSIVE BEHAVIOUR, and TRADE IS MILITARY WARFARE, with EMERGING ECONOMIES ARE A THREAT, IMPROVEMENT IS A CHALLENGE, and EXPANSION IS INVASION. The concordances in Figure 3, featuring the lemmas *conquistare* [to conquer], *vincere* [win], *battaglia* [battle] and *invadere* [to invade], give a taste of these metaphor themes.

The metaphor themes identified all interact with the keywords in ways which can be considered significant. Some of the source domain lexis forms stable collocations with particular key words, such as *strumenti di difesa* [defensive tools] with *commerciale*, *conquistare* [to conquer] with *mercati*, and *sfida* [challenge] with *globalizzazione* [globalisation], and in so doing contribute to the discourse

<p>DELL'IMPORT Il made in Italy argini di crescita passano per la la distribuzione. La necessità di produttive per poi lanciarsi alla ese italiane che hanno tentato di opa è stata una scommessa davvero liana. Sui mercati mondiali siamo namente dimostrano quanto sia più isulta a volte indispensabile per l nostro territorio. Uniti si può utori stranieri: divisi possiamo portunità di lavorare insieme per concorrenza e trovare formule per : divisi possiamo vincere qualche ngo. Diritti e libertà: di tante o di servizi alle imprese, nella conclusa una delle più importanti sono stati visti come pericolo di el tessile, perché è vero che ci uote ma è altrettanto vero che ci</p>	<p>conquista nuovi mercati, Russia e Cina, ma pr conquista dei mercati internazionali. Che cosa conquistare mercati stranieri può offrire nuove conquista dei mercati esteri. Altri paesi hanno conquistare questa fascia alta di mercato vincente anche dal punto di vista degli scambi vincenti solo se riusciamo a trasformare la nos vincente una strategia positiva di adattamento vincere l'agguerrita concorrenza; sostegno che vincere perché siamo in grado di promuovere e v vincere qualche battaglia, ma alla lunga non s vincere la sfida dell'internazionalizzazione. vincerla è necessaria un'azione sinergica che c battaglia, ma alla lunga non sapremo valorizza battaglie per l'affermazione dei diritti indivi battaglia per affermare l'Italia sui mercati battaglie per la supremazia mondiale sul mercat invasione dai cui difendersi". Cina e India ha invaso dopo la fine delle quote ma è altrett ha invaso con un' azione predatoria puntando s</p>
---	--

Figure 3. The emergence of metaphor themes from concordances

norms. However, although conventional and partially delexical, the metaphorical meaning still permeates the discourse:

Through multiple on-line events, certain linguistic forms evolve to become the preferred ways of expressing metaphorical ideas across discourse communities. The language and the conceptual content stabilise, together and co-adaptively, into a particular restricted set of forms and ideas that become part of the resources of language and thinking available in the discourse community.

(Cameron & Deignan 2006: 680)

The institutionalised co-selection of a war metaphor with a key-word makes it extremely difficult for the subject to be mentioned other than in terms of warfare, and this colours perceptions of international trade and commerce in general.

Not all source domain lexis forms such visible collocational relationships with the key words. In this case, the significance of the key metaphors is derived not from the fact that they are significant collocates of the keywords by *upward collocation* (Sinclair 1991: 116), but because the keywords are significant collocates of the source domain lexis, by *downward collocation* (ibid.). Thus a concordance of *Cina* [China] or *India* does not find war-related lexis to collocate particularly frequently, but individual concordances of each of the war-related lexis repeatedly

feature the target domain group including *Cina, India, Asia*, and (*Estremo*) *Oriente* [(Far) East]. Examples 4–6 illustrate this phenomenon.

- (4) ...*ha fatto emergere quel Paese così determinante per il futuro dell'economia mondiale che si chiama Cina. Un Paese che non dobbiamo temere...*
 ...allowed that country which will decide the future of the world economy to emerge: China. A country which we must not fear...
- (5) *La paura che le nostre imprese uscissero soccombenti dalla sfida della globalizzazione, soprattutto di fronte all'aggressività commerciale dell'Estremo Oriente...*
 The fear that our businesses will end up as the loser in the globalisation challenge, especially when faced with the commercial aggressiveness of the Far East...
- (6) *Ma la Cina rappresenta il vero futuro del tessile, perché è vero che ci ha invaso dopo la fine delle quote ma è altrettanto vero che ci ha invaso con un'azione predatoria puntando sulla bassa qualità e sul basso prezzo.*
 'But China is the real future of the textile [industry], because it is true that it has invaded us since quotas ended but it is just as true that it has invaded us with predatory behaviour aiming at low quality and low prices.'

This kind of co-selection is important because it is hidden, and because it is pervasive. If most instances of the lemma *invadere* have China, India, Asia, or the Far East as collocates, then the meaning of *invadere* comes to be associated with those nations, as the "use and re-use of metaphors leads to the conventionalization of attitudinal judgements attached to them" (Cameron & Deignan 2006: 676). The relationship between the target and the source evades identification because downward collocation, while providing semantic information about a word (Sinclair 1991: 116), concerns low-frequency lexis which is rarely the focus of any linguistic enquiry, with or without the aid of corpora. The war metaphor is key, but covertly so.

5. Conclusions

The main difference between the keyness represented by statistically-generated key words and that represented by the metaphors discussed above, is that the former is overt while the latter is hidden from view. Overt keyness applies to those words that are frequent enough and prominent enough to attract attention. It sums up the aboutness of the text or discourse, insofar as its topics and themes are concerned. Covert keyness, on the other hand, can be hidden from statisti-

cal measures of significance because it describes repetition which operates above the level of the word. The abstract nature of covert keyness can make it difficult to pinpoint, but once located it sheds light on the underlying attitudinal stance expressed in the texts. Thus overt keyness tells us *what* is key, and covert keyness tell us the reasons *why*.

Keywords are neutral, but key metaphors are not. The war metaphors discussed in 4 are not limited to the Minister's speeches, but are found in the language at large, feeding the perception that emerging markets pose a threat to Italy, that they must be fought, and that the national territory must be defended at all costs. The textual use and function of metaphor themes interact with the aboutness of the text, and they do so in ways that are perceptible, yet barely visible. Downward collocation is difficult to spot without the use of concordancing software, but once identified, it can provide substantiation for those perceptions. In this data, the key word list indicates that China (and latterly, India⁹) is important, but does not provide the reasons for the prominence given to this nation over any of the others in the corpus. By examining the metaphorical activity in the data, however, and by examining the interaction of the covert and overt keyness, two aspects of Italo-Chinese relations come to the surface: (i) that Chinese exports and current Chinese economic policy are perceived as a threat to the Italian economy, and (ii) that China is mentioned frequently because of the threat it poses. Thus within this discourse, the key word *Cina* is seen to be bound up with the key metaphor theme of war; and pragmatically, China is painted as an enemy and a threat.

The evaluative function of metaphor is well documented, but rarely refers to anything other than instances of linguistic metaphors in texts. It has been demonstrated here that metaphor themes also play an important evaluative role, both in text and within specialised discourse as a whole. It has also been demonstrated, however, that the source domain lexical items are not interchangeable. In common with all words, metaphors form units of meaning, and it is by examining the collocational profiles of each source domain lexical item that the metaphor's evaluative meaning is built up. Abstract metaphor themes make it possible to make generalisations about key lexis, but the details are found in collocation.

9. The appearance of India on the key word list is a more recent phenomenon; it was not present in the list generated during a preliminary study carried out on the data spanning June 2006 – January 2007.

References

- Black, M. 1993. More about metaphor. In Ortony (ed.), 19–41.
- Cameron, L. & Deignan, A. 2006. The emergence of metaphor in discourse. *Applied Linguistics* 27(4): 671–690.
- Cameron, L. & Stelma, J. 2004. Metaphor clusters in discourse. *Journal of Applied Linguistics* 1(2): 107–136.
- Charteris-Black, J. 2004. *Corpus Approaches to Critical Metaphor Analysis*. Basingstoke: Palgrave Macmillan.
- Charteris-Black, J. 2005. *Politicians and Rhetoric: The Persuasive Power of Metaphor*. Basingstoke: Palgrave-MacMillan.
- Deignan, A. 2005. *Metaphor and Corpus Linguistics* [Converging Evidence in Language and Communication Research 6]. Amsterdam: John Benjamins.
- Deignan, A. 1999. Linguistic metaphors and collocation in non-literary corpus data. *Metaphor and Symbol* 14(1): 19–36.
- Glucksberg, S. & Keysar, B. 1993. How metaphors work. In Ortony (ed.), 401–424.
- Jäkel, Ö. 1999. Kant, Blumenberg, Weinrich: Some forgotten contributions to the cognitive theory of metaphor. In *Metaphor in Cognitive Linguistics* [Current Issues in Linguistic Theory 175], R. W. Gibbs & G. J. Steen (eds), 9–27. Amsterdam: John Benjamins.
- Koller, V. & Semino, E. 2009. Metaphor, politics and gender: A case study from Germany. In *Politics, Gender, and Conceptual Metaphors*, K. Ahrens (ed.), 9–35. Basingstoke: Palgrave Macmillan.
- Lakoff, G. & Johnson, M. 1980. *Metaphors We Live By*. Chicago IL: The University of Chicago Press.
- Louw, W. E. 2000. Some implications of progressive delexicalisation and semantic prosodies for Hallidayan metaphorical modes of expression and Lakoffian “Metaphors We Live By”. Privately-distributed version of: Progressive delexicalization and semantic prosodies as early empirical indicators of the death of metaphors. Paper read at the 11th Euro-International Systemic Functional Workshop: Metaphor in systemic functional perspectives, University of Gent (Belgium), 14–17 July 1999.
- Ortony, A. (ed.). 1993. *Metaphor and Thought*, 2nd & rev. edn. Cambridge: CUP.
- Partington, A. 2003. *The Linguistics of Political Argument: The Spin-doctor and the Wolf-pack at the White House*. London: Routledge.
- Pragglejaz Group. 2007. MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol* 22(1): 1–39.
- Philip, G. 2009. ‘Non una donna in politica, ma una donna politica’: Women’s political language in an Italian context. In *Politics, Gender, and Conceptual Metaphors*, K. Ahrens (ed.), 83–111. Basingstoke: Palgrave Macmillan.
- Philip, G. In press a. *Colouring Meaning: Collocation and connotation in figurative language* [Studies in Corpus Linguistics]. Amsterdam: John Benjamins.
- Philip, G. In press b. Locating metaphor candidates in specialised corpora using raw frequency and key-word lists. In *Metaphor in Use: Context, Culture, and Communication*, F. MacArthur, J. L. Oncins-Martínez, M. Sánchez-García & A. M. Piquer-Píriz. Amsterdam: John Benjamins.
- Rayson, P. 2005. *Wmatrix: A Web-based Corpus Processing Environment*. Computing Department, Lancaster University. <<http://www.comp.lancs.ac.uk/ucrel/wmatrix>>.

- Richards, I. A. 1936. Metaphor. In *The Philosophy of Rhetoric*, I. A. Richards (ed.). London: OUP.
- Scott, M. 2004. *WordSmith Tools*, Version 4. Oxford: OUP.
- Sinclair, J. M. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP.
- Sinclair, J. M. 1996. The search for units of meaning. *Textus* 9(1): 71–106.
- Steen, G. 1994. *Understanding Metaphor in Literature*. London: Longman.
- Weinrich, H. 1958. Münze und Wort: Untersuchungen an einem Bildfeld. In *Sprache in Texten*, H. Weinrich (ed. 1976), 276–290. Stuttgart: Klett.
- Weinrich, H. 1967. Allgemeine Semantik der Metapher. In *Sprache in Texten*, H. Weinrich (ed. 1976), 317–327. Stuttgart: Klett.

Appendix 1

Top 50 content words. Keywords appear in bold.

1. ITALIA [Italy]
2. IMPRESE [businesses]
3. MINISTRO [minister]
4. PAESI [countries]
5. COMMERCIO [commerce]
6. INTERNAZIONALE [international]
7. PAESE [country]
8. CRESCITA [growth]
9. MERCATI [markets]
10. GOVERNO [government]
11. SISTEMA [system]
12. CINA [China]
13. POLITICHE [political]
14. SVILUPPO [development]
15. COMMERCIALE [commercial]
16. MERCATO [market]
17. SETTORE [sector]
18. EUROPEA [European]
19. PARTE [part]
20. COMMISSIONE [commission]
21. EUROPA [Europe]
22. INTERNAZIONALIZZAZIONE [internationalisation]
23. ITALY –
24. ECONOMIA [economy]
25. EUROPEE [European]
26. PRODOTTI [products]
27. POLITICA [politics]
28. INDIA [India]
29. ESTERO [foreign]
30. MINISTERO [ministry]
31. MONDO [world]

32. ANNI [years]
33. ESPORTAZIONI [exports]
34. PARTICOLARE [particular]
35. INVESTIMENTI [investments]
36. ITALIANO [Italian]
37. MADE –
38. ITALIANE [Italian]
39. EURO [Euro]
40. ECONOMICO [economic]
41. STATI [states]
42. ANNO [year]
43. ECONOMICA [economic]
44. SETTORI [sectors]
45. UE [E.U.]
46. PRESIDENTE [president]
47. MAGGIORE [greater]
48. RISPETTO [(with) respect (to)]
49. EUROPEO [European]
50. ITALIANA [Italian]

Appendix 2

Key words

1. AREA [area]
2. AZIENDE [companies]
3. CINA [China]
4. COMMERCIALE/I* [commercial]
5. COMMERCIO [commerce]
6. CONCORRENZA [competition]
7. CRESCITA [growth]
8. ECONOMIA [economy]
9. ECONOMICO/A [economic]
10. EMERGENTI [emerging]
11. ESPORTAZIONI [exports]
12. ESTERO/I [foreign]
13. EUROPEA/E [European]
14. GLOBALIZZAZIONE [globalisation]
15. IMPORTAZIONI [imports]
16. IMPRENDITORIALE [entrepreneurial]
17. IMPRESE [businesses]
18. INDIA [India]
19. INDUSTRIA [industry]
20. INDUSTRIALE/I [industrial]
21. INFRASTRUTTURE [infrastructure]
22. INFRAZIONE [infringement]

23. INTERNAZIONALE [international]
24. INTERNAZIONALIZZAZIONE [internationalisation]
25. INTERSCAMBIO [interchange]
26. INVESTIMENTI [investments]
27. ITALIA [Italy]
28. ITALY –
29. MADE –
30. MANIERA [manner/ way]
31. MEDIE [medium-sized]
32. MERCATO/I [market(s)]
33. MILIARDI [billion]
34. MONDIALE [world]
35. PICCOLE [small]
36. PROCEDURA/E [procedure(s)]
37. PRODOTTI [products]
38. PRODUTTIVO [productive]
39. PRODUZIONE [production]
40. RIPRESA [upturn]
41. SETTORE/I [sector(s)]
42. SISTEMA [system]
43. STRATEGIA [strategy]
44. SVILUPPO [development]
45. UE [E.U.]

* Alternative forms are provided only when they appear on the key-word list; in all cases, the gender and number of the listed word forms is to be considered significant.

Keyness in Texts

Edited by

Marina Bondi

University of Modena & Reggio Emilia

Mike Scott

Aston University

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Cover design: Françoise Berserik
Cover illustration from original painting *Random Order*
by Lorenzo Pezzatini, Florence, 1996.

Library of Congress Cataloging-in-Publication Data

Keyness in texts / edited by Marina Bondi and Mike Scott.

p. cm. (Studies in Corpus Linguistics, ISSN 1388-0373 ; v. 41)

Includes bibliographical references and index.

1. Semantics.
 2. Discourse analysis.
 3. Corpora (Linguistics)
 4. Phraseology.
- I. Bondi, Marina. II. Scott, Mike, 1946-

P302.K48 2010

401'.43--dc22

2010029119

ISBN 978 90 272 2317 3 (Hb ; alk. paper)

ISBN 978 90 272 8766 3 (Eb)

© 2010 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Perspectives on keywords and keyness: An introduction <i>Marina Bondi</i>	1
Section I. Exploring keyness	
Three concepts of keywords <i>Michael Stubbs</i>	21
Problems in investigating keyness, or clearing the undergrowth and marking out trails... <i>Mike Scott</i>	43
Closed-class keywords and corpus-driven discourse analysis <i>Nicholas Groom</i>	59
Hyperlinks: Keywords or key words? <i>Jukka Tyrkkö</i>	79
Web Semantics vs the Semantic Web? The problem of keyness <i>François Rastier</i>	93
Section II. Keyness in specialised discourse	
Identifying aboutgrams in engineering texts <i>Martin Warren</i>	113
Keywords and phrases in political speeches <i>Denise Milizia</i>	127
Key words and key phrases in a corpus of travel writing: From Early Modern English literature to contemporary “blooks” <i>Andrea Gerbig</i>	147
History v. marketing: Keywords as a clue to disciplinary epistemology <i>Donatella Malavasi and Davide Mazzi</i>	169

Metaphorical keyness in specialised corpora <i>Gill Philip</i>	185
Section III. Critical and educational perspectives	
A contrastive analysis of keywords in newspaper articles on the "Kyoto Protocol" <i>Erica Bassi</i>	207
Keywords in Korean national consciousness: A corpus-based analysis of school textbooks <i>Soon Hee Fraysse-Kim</i>	219
General spoken language and school language: Key words and discourse patterns in history textbooks <i>Paola Leone</i>	235
Index	249