Perspectives on keywords and keyness
An introduction

Marina Bondi
University of Modena and Reggio Emilia, Italy

> "All words are equal, but some are more equal than others"
> (adapted from Orwell's *Animal Farm*)

Lexical items enjoy equal status in the lexicon of a given language, but their importance varies from the point of view of text. Each individual word form contributes to the construction of meaning in text, but only some words are key-words, i.e. words that play a role in identifying important elements of the text. Similarly, any given language is constituted by all the lexical elements that become part of it, but only some lexical elements are taken to characterize its cultural specificity.

Starting from the different interpretations of the expression "keywords" – as searching tools, in text mining and classification, but also as analytic tools in text interpretation and discourse analysis – this introduction focuses on the relationship between words and text, looking at the co-text of the word, but also at the cultural context that informs the text, where culture is taken to mean the repertoires of meanings shared within a community (e.g. national, or local, but also disciplinary). Keywords are often taken to be markers of the "aboutness" and the style of a text (Scott & Tribble 2006:59–60): what we want to investigate here is what structures of textuality keywords point to and how far they are also influenced by the position of the writer, in the context of text production.

1.   Keywords and keyness in language studies

The notion of keyword has no well-defined meaning in language studies. The definition of a "word" as such may be seen as problematic in modern linguistics; de Saussure's search for the basis of a scientific study of language as system led him to units different from the word at various levels of analysis – phonetics and phonology, syntax, morphology, semantics.

Lexical analysis has long been concerned with the ways in which language, and lexis in particular, instantiates culture. Already in the nineteen-thirties, Firth's lexical semantics proposed the study of "sociologically important words," what one might call focal or pivotal words" and advocated an analysis of the distribution of words whose meanings characterize a community by occurring in specific contexts, with specific associations and values (1935:40–41). On the basis of anthropological notions of context, and referring in particular to Malinowski, his colleague at the London School of Economics, Firth showed how the study of words in context can illuminate meanings that characterize a culture and a community, referring for example to the development of the meanings of *clerk* in Middle English from medieval clerics.

Similarly, Cultural Studies – Williams (1976) in particular – made an attempt to produce an analysis of contemporary culture through the study of a number of "cultural keywords," i.e. the 'dictionary' of a culture and a social group. The meanings of words like *alienation, capitalism, family, fiction, hegemony, literature, media, tradition* etc. were taken to represent the most distinctive features of contemporary western culture, by integrating synchronic and diachronic perspectives in a full appreciation of meaning. Williams thus made the link between keywords and discourse communities even more explicit, but he clearly oriented the analysis to historical and social macro-contextual factors only, not paying much attention to text and genre and leaving methodological tools for the analysis of meaning completely undiscussed.

A similar focus, but with a different perspective – oriented to the distinction between semantic universals and cultural underpinnings of a language – is provided by Anna Wierzbicka (1999, 2006). Wierzbicka looks at lexical semantics through her Natural Semantic Metalanguage (NSM) as a key to the history, culture and society that produced it, considering the impact of values on interaction and its strategies. Her approach aims at counteracting both the tendency to mistake Anglo English for the human norm and widespread attempts to deny the existence and continued relevance of the cultural baggage of English in international communication. She looks for example at typical features of "anglo" culture such as the ideal of accuracy, the practice of understatement, recourse to "facts" and emphasis on rationality as against emotions. The importance of the meanings associated with a word like *reasonable* shows that "reasonableness" may prove to be the most effective persuasive strategy in an anglophone cultural context, which leaves little room to asymmetrical relations and denies persuasive power to both pleading and authority claims. Her study of the historically shaped cultural meanings of words like *right, wrong, reasonable, fair* aims at revealing covert meanings making the heritage of a common spirit perceivable. Her framework combines cognitive and interactional perspectives, attention to thinking, speaking and

doing, with an interesting emphasis on the impact of values on interactive strategies, while still relying on an intuitive process of keyword and data selection.

Keywords are not necessarily a key to culture, however: they may facilitate understanding of the main point of a text, constituting chains of repetition in text. Whether referring to words that are key to the intepretation of a text or key to the interpretation of a culture, the study of keywords has become central in corpus linguistics, especially through the development of techniques for the analysis of the meaning of words in context. In a quantitative perspective, keywords are those whose frequency (or infrequency) in a text or corpus is statistically significant, when compared to the standards set by a reference corpus (Scott 1997; Baker 2004; Scott & Tribble 2006).

Identifying elements that are repeated to a statistically significant extent does not in itself constitute an analysis or an interpretation of the text or corpus. It does however point to elements that may be profitably studied and need to be explained. It certainly does point to fundamental elements in describing specialised discourse or in placing a text in a specific domain. The problem for the researcher, of course, lies both in the design of appropriate and adequately representative corpora and in the delicacy of the analysis, with its capacity to isolate specific questions and avoid overgeneralization.

In a corpus perspective, keywords are studied through their typical co-occurrence with other lexico-semantic units. Michael Stubbs (1996, 2001), for example, has shown the importance of concordance analysis in this field: the cultural and ideological implications of a lexical element can be illuminated by an analysis of its collocation and semantic preference – the tendency of the word to co-occur with other words and with words belonging to a specific semantic category or field (see also Sinclair 1996).

The notion of quantitative keyness applies equally to word forms, lemmas and word sequences[1]. The definition thus easily adapts to more complex units than the word, pointing towards a perspective that is gaining ground in present-day descriptive and theoretical language studies: phraseology. Keywords, in fact, are not necessarily single words: we can look at key-clusters (repeated strings of words)[2] or even key-phrases, when extended units of meaning (Sinclair 1996)

---

1. In a semantic perspective the notion has also been recently extended to semantic elements (Rayson 2008). These of course can only be based on previous semantic analysis and tagging of the corpus, on the basis of given semantic descriptors.

2. In the field of *natural language processing*, computational linguistics and corpus linguistics, research on 'n-grams,' also called 'word clusters,' 'lexical clusters' or 'bundles' (cf. Biber, Conrad & Cortes 2004; Carter & McCarthy 2006) studies contiguous word forms building up to create repeated word sequences in the corpus.

are considered, i.e. words in combination originating a unit of meaning that can be different from the sum of the constituent lexical units. In the words of John Sinclair (2005), a corpus perspective looks at words in combination and finds in phraseology the ideal starting point for the exploration of the systematic relation between text and form.

Emphasis on phraseology has been increasing in corpus research (e.g. Hunston & Francis 2000; Moon 2002; Hunston 2004). The revived interest finds its origin in Sinclair's notion of collocation (e.g. Sinclair 1991) and in his "idiom principle", highlighting that in the linearity of text each choice narrows down the range of possible choices in the elements that follow and that "a language user has available to him or her a large number of semi-preconstructed phrases that constute single choices" (1991:110).

Phraseological studies have shown a tendency to shift their attention from fixed, opaque multiword units to a much wider range of units. The focus of interest can thus be extended to discontinuous or inverse relations ("concgrams", Cheng, Greaves & Warren 2006) and patterns (Hunston & Francis 2000).

The key lexical elements of a text create a dense network of intercollocation, including both continuous and discontinuous phraseological patterns. Following Phillips (1989), for example, we can look at the patterns created in text between *electric* and *charge*, but also at the patterns created in text between their collocates (e.g. for *charge: distribution, density, point, uniform*; for *electric: dipole*). The network of lexical relations of this kind would contribute to an identification of the "aboutness" of a text.

When lexical analysis combines with semantic analysis, looking at the extended unit of meaning with its corollary of semantic preference and semantic prosody (Sinclair 1996), attention to the co-text means identifying both the potential semantic associations between otherwise different forms and the association of the unit with further textual-pragmatic meanings. A recent development along these lines is the corpus study of semantic sequences, i.e. "recurring sequences of words and phrases that may be very diverse in form [...] more usefully characterised as sequences of meaning rather than as formal sequences (Hunston 2008:271).

2.   **The keyness metaphor in knowledge management: "Aboutness" as subject matter**

The meaning of keywords is often explored through the metaphor on which the expression is based. A key is a tool that gives you access to something. The metaphor refers to the power of opening (and closing), revealing (or hiding) what is unknown or unclear. A keyword gives access to features of a text or corpus that

are not immediately obvious: but what are these features? What textual doors are opened by keywords?

The first meaning of *keyword* is perhaps the most obvious in knowledge management, where keywords are those that help identify a text in structured data-bases, such as for example library resources. Textual data-bases can be searched making use of keywords to be automatically retrieved in pre-defined fields: title, author, abstract, subject descriptors, or the text itself. A range of tools is needed because titles are not always the best indicators of the subject matter of a text. This is quite obvious in literary writing: no-one thinks of *Hamlet* or *Othello* as indicators of subject matter or theme. It is less obvious but equally true of professional communication. A text entitled *The Danger Model: A Renewed Sense of Self* cannot be automatically attributed to a subject or a discipline. When we see it is a viewpoint article published in *Science* in 2002, we can probably exclude some of the expectations created by the title, but we need at least descriptors to understand that the field is immunology. The abstract reveals that the text discusses a change in paradigm in immunological studies, a shift from a vision of immunology as based on the distinction between self and non-self to a vision of the immunological system as worried about danger rather than foreignness. Titles can be seen as a key to texts, though not always the most direct key to their subject matter.

With the proliferation of scientific publications and the ever increasing use of textual data-bases, keyword searches have become central to knowledge management. Subject classification, however, is mostly realized from a perspective that is external to the text itself, making use of bibliographical classifications of knowledge such as the Dewey system. Author-produced keywords have also been used, though with unstable results. A priori categorizations are intersubjectively valid but they lack flexibility. Author-produced keywords are more flexible but they lack intersubjective comparability. Knowledge representation has become a key issue: from general and domain ontologies, to semantic networks and "frames". The attention often shifts from lexical units characterizing the surface of text to the possibility of recovering meaning structures beyond lexical forms.

The development of information science and of web-based knowledge, however, has shifted attention from information retrieval to information extraction. The availability of enormous quantities of unstructured data on the web poses the question of information "extraction": text mining requires tools that can move from lexical forms to meanings and their structures, thus finding keywords through the text rather than outside the text. In text mining, just like in current linguistic research, phraseological units are gaining importance. Tools for the identification of keywords are being developed on the basis of frequency data that do not simply look at individual word forms, but rather at relations between words that frequently co-occur.

This brings us back to keywords as words whose frequency (or infrequency) in a text or corpus is statistically significant. The vast majority of the keywords that can be determined by automatic analysis of a text will be key to its subject matter.

3. The keyness metaphor and text interpretation: Subject matter and organization

The notion of text has been one of the most influential in theoretical and methodological developments in linguistics. The past forty years have shown growing interest in meaning making processes beyond the basic syntactic unit of the clause or clause complex, starting for example with work on textuality and text cohesion (e.g. Halliday & Hasan 1976; Beaugrande & Dressler 1981; Conte, Petöfi & Sözer eds. 1989) and leading up to recent interest in patterns of lexis in text (Hoey 1991) as well as meaning units in the linearity of text (Sinclair 2004; Sinclair & Mauranen 2006). Lexical elements can be shown to play a key role in the cohesion of text (signalling and establishing relations between lexical units) and in textual coherence (the conceptual and functional unity of a text). In such a textual perspective, words can become key to the conceptual structure of the text – very much in the same way as in librarianship they define its subject matter – but also to the organizational structure of text – in ways that may also be illustrative of its communicative purpose.

Cognitive and pedagogic approaches to text have often shown that for the act of reading the words that organize text may be more important than those that identify its "content", because they guide the reader towards the elements of content. Signals of organizational structure will thus be key (or "pivotal") in reading because they facilitate access to the information required. If exploring a data-base requires use of keywords that constitute a map of existing knowledge, exploring a text requires use of organizational keywords that act as a textual map. Keywords signalling textual organization act as signposts and help readers identify generic patterns and locate information. When looked at from this perspective, keyness also links to a vast literature on meta-discourse and its role in reading (e.g. Vande Kopple 1985; Crismore 1989; Hyland 2005; Ädel 2006). Let us take the basic metadiscursive structure of two abstracts like the following as an example:

(1a) In this paper we investigate the implications of... In the received theory of ...In our model... Hence, we conclude that ....

(1b) Recent studies highlight increasing recognition of... It is understood now that ..... This article overviews ... and outlines..., including ...

Irrespective of whether we are talking about nanotechnologies or market structures, the elements reported highlight the basic communicative structure of the text they represent: in the first case the text presents a new model contrasting it with more consolidated theories, while the second introduces a critical review of an issue on the background of recent developments in the field. If subject matter is essential in retrieval, communicative purpose and genre (research article vs review article) may be equally important in reading and metadiscursive elements act as signposts to actual content.

Key-words, key-clusters and key-phrases are not always elements of the conceptual structure of a text. There may be elements of grammatical structure or elements of self-reference. These become useful pointers to the most frequent textual structures of a text as well as its most frequent metadiscursive phaseology.

We may thus think of two kinds of keywords, much in the same way as Sinclair and Mauranen's "Linear Unit Grammar" distinguishes two kinds of unit in the linearity of text – "message-oriented elements", contributing to the topical continuation of discourse, and "organization-oriented elements", that contribute to managing discourse (2006:59–60). On the one hand, there are keywords that point at the conceptual structure of a text, its "aboutness", what the text is about. On the other, there are keywords that point at issues that may prove to be useful indicators of the communicative purpose and micro- or macro-structure of the text, what the text does and how.

4. Keyness in text and discourse: A sample analysis

As will be apparent from the rest of the volume, work on keyness in text easily leads to work on discourse, linking language use beyond the sentence to the study of social practices and ideological assumptions associated with language (thus involving the different definitions of discourse listed by Schiffrin, Tannen & Hamilton 2001:1). Words and phrases that are key in a text or in a corpus may be shown to be indicative of the writer's position and identity, as well as of the discourse community with its values and beliefs about the subject matter and the genres that characterize it (e.g. Baker 2006; Biber, Conrad & Cortes 2007; Ädel & Reppen eds. 2008).

In studies of academic discourse, for example, the acquisition of academic literacy has often been seen as a process of enculturation of students into disciplinary communities through a process of informal learning, of apprenticeship into the ways of speaking of the community (Berkenkotter & Huckin 1995:7). Academic discourse communities are seen by John Swales as social groupings identified by "a broadly agreed set of common public goals", participatory mechanisms of intercommunication, specific genres and lexis, and "a threshold level of members

with a suitable degree of relevant content and discoursal expertise" (1990: 24-27). Research perspectives have become more and more interested in cross-disciplinary analysis, focusing on the role played by "disciplinary culture in defining what is conventionally seen as acceptable argument or textual organization" (Hyland 2000; Hyland & Bondi eds. 2006). Cross-disciplinary research has recently extended the attention traditionally paid to domain terminology to include interest in general, particularly in the "general academic lexis" that is used across a wide span of domains. It has been shown that different disciplines tend to use it in slightly different ways, on the basis of their methodological tenets. A word like *case*, for example, is frequently used both in economics and in business studies, but it is used in contexts that are fundamentally different and representative of different argumentative frameworks. The word occurs most frequently in collocations like *(the) case of* or *(BE) the case* in economics, thus signalling the setting up of hypotheses and scenarios, whereas in business studies it is more often found in expressions like *case study* or *case in point*, signalling an exemplification or an illustration (Bondi 2006).

The words and expressions that recurrently identify the conceptual structures and the organizational structures of a text or corpus can be studied to illuminate features of the discourse that produces the text or corpus. The keywords that point to the aboutness of a text or corpus will be key to the ontology of the discourse. The keywords that point to textual organization will be key to the epistemology.

We can explore these preliminary statements through a case study of a landmark text: the *General Theory of Employment, Interest and Money* by John Maynard Keynes.3 In the full title – *The General Theory of Employment, Interest and Money* – we can see that *Employment, Interest and Money* have been chosen as key to the subject matter, whereas the choice of *General Theory* is meant to provide a form of self-representation that highlights the main communicative purpose of the writer, as well as his theoretical position. In the first one-paragraph chapter of the volume, Keynes presents his position against classical economic theories, anticipating a criticism of their fundamental postulates as based on a special case, rather than a more general vision:

(2) Chapter 1 – THE GENERAL THEORY

I have called this book *the General Theory of Employment, Interest and Money*, placing the emphasis on the prefix *general*. The object of such a title is to contrast the character of my arguments and conclusions with those of

3. John Maynard Keynes, *The General Theory of Employment, Interest and Money*, New York, Harcourt and Brace 1936; e-text available from The University of Adelaide Library Electronic Texts Collection (http://ebooks.adelaide.edu.au/).

the classical theory of the subject, upon which I was brought up and which dominates the economic thought, both practical and theoretical, of the governing and academic classes of this generation, as it has for a hundred years past. I shall argue that the postulates of the classical theory are applicable to a special case only and not to the general case, the situation which it assumes being a limiting point of the possible positions of equilibrium. Moreover, the characteristics of the special case assumed by the classical theory happen not to be those of the economic society in which we actually live, with the result that its teaching is misleading and disastrous if we attempt to apply it to the facts of experience.

The contrast between *special* and *general* provides the starting point for the whole book. The book itself is commonly called "The General Theory", thus giving prominence to what Keynes presented as the element of novelty of his book.

An analysis of the keywords of the text will clearly show that the words in the title also recur as keywords. Using *Wordsmith 5* (Scott 2008), we have calculated keywords with reference to different corpora: a previous book by the same author (*The economic consequences of the peace4*), a reference corpus of current economic articles (HEM-Economics5) and a general reference corpus (BNC-written component).

The relative positions of *employment, interest* and *money* vary slightly but they remain among the top five keywords in all three cases. Although not too much weight can be placed on the order of KWs, as argued by Scott (this volume), nevertheless where the terms are of similar frequency the first positions in keyword lists are often indicative of subject matter and they are relatively consistent across corpora.

The other words included in the top 5 are also worth considering. Both the general written language and Keynes' previous book highlight other content words: *rate* and *investment*. These point at important conceptual elements of the *General Theory* that distinguish it from previous work. The word *rate* is typically used in the cluster *the rate of interest* (348/737 occurrences), which is one of the foci of

4. John Maynard Keynes, *The Economic Consequences of the Peace*, New York, Harcourt, Brace and Howe, 1920; E-text available from The Project Gutenberg Online (http://www.gutenberg.org/files/15776/15776.txt).

5. The corpus comprises 436 articles published in 2000-2001 from the following journals: *European Economic Review* (EER), *European Journal of Political Economy* (EJOPE), *International Journal of Industrial Organization* (IJOIO), *International Review of Economics and Finance* (IREF), *Journal of Corporate Finance* (IJOCF), *Journal of Development Economic* (IODE), *Journal of Socio-Economics* (IOSE), *The North American Journal of Economics and Finance* (NAJEF).

the book, essential to a theory of investment. *Investment* also marks an important shift in Keynes' work, moving from the international economic framework of the first book – basically an example of economic historical analysis – to an emphasis on the micro-economic foundations of macro-economics in the *Theory*.

The corpus of current economics articles, on the other hand, highlights grammatical words: *of* and *which*. *Which* is mainly to be attributed to frequent use of relative defining and non-defining clauses; a look at collocates will show that the nouns specified by the relative include all the important conceptual elements of the text. Here are the top twenty nouns in the position immediately preceding the relative, in order of decreasing frequency: *interest, factors, investment, money, employment, equipment, level, amount, income, consumption, factor, rate, capital, cash, sum, production, cost, theory, demand, value*. The list includes the three keywords we started from, as well as many other words referring to related concepts.

The presence of *of* reflects a preference for nominal postmodification against noun + noun constructions. If we look at the clusters it is found in, we see a vast dominance of important phraseological structures: *the marginal efficiency of capital, the rate of interest, the quantity of money* are the most frequent, variously related to a theory of investment.

The grammatical words we have found, then, do not point directly at the subject matter of the text, but rather at typical constructions used: the complex nominals used to characterize complex notions and the need to define these in terms of the processes they are characterized by. Both *which* and *of* can be seen as matters of individual style. In this case, however, they are more likely to reflect differences in register due to genre and diachronic change. The reference corpus in fact is representative of a much denser form of writing and also of a much later stage in the development of the discipline, a stage in which terminology based on nominal contructions has been developed to a great extent.

Moving from simple keywords to key-clusters we are more likely to find the complex notions expressed in phraseological terms and also to find other pointers to the typical structure of discourse. Looking at 3–5-word clusters with reference to the *Economic Consequences of the Peace* and to the current economics articles shows very similar clusters, with 4/5-word content keyphrases like *the rate of investment, (the) marginal efficiency of capital* and *the quantity of money* among the top five. 3-word lists also show a number of organizational key-phrases like: *in terms of, is equal to, as a whole, it is the, in the sense, it follows that, as a rule* etc. While hardly signals of subject matter, all these expressions act as signals of frequent communicative acts: defining (*in terms of, in the sense*), identifying in mathematical terms (*is equal to*), highlighting (*it is the*), deducing (*it follows that*) etc. These can again be considered matters of individual style, but they also point at important features of the genre and of the writer's authorial identity, highlighting that

we are dealing with scientific argument that favours logical structures. The fact that signals of logical deduction are particularly distinctive in comparison with Keynes' earlier work and lose keyness in comparison with current economics articles may be taken as a sign of authorial development towards forms of more formal reasoning that will become characteristic of the discipline at later stages.

Keywords can also be calculated for each chapter with reference to the whole book. Chapter 18, for example, stands out as characterized by relatively little specific language and rather an insistence on general academic lexis, with keywords like *factors, variables, condition, psychological, we*. The most likely explanation of this peculiarity seems to me to lie in the summative nature of the chapter, which is entitled "The General Theory of Employment Re-stated" and begins by stating: "We have now reached a point where we can gather together the threads of our argument". At other points, negative keywords – words that stand out as particularly infrequent – will play an equally important role: *money*, for example becomes a negative keyword at regular points in the book – Chapter 6, 8, 22 and 24, where Keynes tries to correct monetary views of income, saving and investment, analyses the propensity to consume, explains the trade cycle and sums up his social philosophy.

Similarly, if we look at the well known 1936 Preface of the book, we get a very simple picture, with only three keywords.

N	Keyword	Freq.	%	RC. Freq.	RC. %	Keyness	P
1	I	25	2.39	457	0.40	46.83	0.000000000
2	MY	13	1.24	120	0.11	39.21	0.000000000
3	BOOK	7	0.67	50	0.04	24.20	0.0000008665

Figure 1. Keywords of the Preface of the General Theory

The Preface is, entirely predictably, about the writer and his book. It is interesting, however, to check the concordances and see that all the three keywords are in fact self-reference items in the Preface. The choice between personal and non-personal elements of self-reference can be better studied in the co-text of concordance lines and a tendency can be noticed to use non-personal reference to introduce the most important statements about the nature of the book that follows, including some among the best known quotes from the text. The occurrences mark the main steps in the rhetorical organization of the preface:

a. Definition of the intended audience: *This book is chiefly addressed to my fellow economists. I hope that it will be intelligible to others. But its main purpose is to deal with difficult questions of theory, and only in the second place with the applications of this theory to practice.*

b. Relation to previous work on monetary policy: *The relation between this book and my Treatise on Money [JMK vols. v and vi], which I published five years ago, is probably clearer to myself than it will be to others; and what in my own mind is a natural evolution in a line of thought which I have been pursuing for several years, may sometimes strike the reader as a confusing change of view.[...] This book, on the other hand, has evolved into what is primarily a study of the forces which determine changes in the scale of output and employment as a whole; and, whilst it is found that money enters into the economic scheme in an essential and peculiar manner, technical monetary detail falls into the background.*

c. Acknowledgement of colleagues' support: *The writer of a book such as this, treading along unfamiliar paths, is extremely dependent on criticism and conversation if he is to avoid an undue proportion of mistakes. [...] In this book, even more perhaps than in writing my Treatise on Money, I have depended on the constant advice and constructive criticism of Mr R. F. Kahn. There is a great deal in this book which would not have taken the shape it has except at his suggestion.*

d. Definition of the innovative nature of the book: *The composition of this book has been for the author a long struggle of escape, and so must the reading of it be for most readers if the author's assault upon them is to be successful, – a struggle of escape from habitual modes of thought and expression. The ideas which are here expressed so laboriously are extremely simple and should be obvious. The difficulty lies, not in the new ideas, but in escaping from the old ones, which ramify, for those brought up as most of us have been, into every corner of our minds.*

To conclude, we can focus on the word *theory*, which we saw so prominent in the title and in Chapter 1. *Theory* is again an important keyword with all the three reference corpora (with a keyness score of 230.05 when measured against Keynes's previous work, 232.59 against current economics articles, 881.6 against general writing). The word *general*, on the other hand, is not key in comparison with Keynes' previous work and its keyness score in comparison with current economic writing is relatively low (31,32). It is true however that if we look at the collocation *general theory*, then this becomes highly distinctive: the expression is absent from earlier work and it becomes virtually exclusive of Keynes in current economic writing (49 of the 50 occurrences in the corpus of economics articles are references to Keynes' book).

The contrast between the classical theory and Keynes' own General Theory is emphasized by the fact that the latter is mostly presented as what we assume (as against what the classical theory assumes), with an intensive use of the first

General Theory as such is mostly present in sections and chapters that are characterized by important metadiscursive nodes – introductory sections and conclusion – whereas in the development of the argument the controversy is between "us" and the classical theorists.

The word *classical*, on the other hand, is key everywhere, though most prominent in comparison with general writing (keyness score 514.3), followed by current articles (240.64) and Keynes' previous work (114.61). Keynes admittedly devoted much of his book to a refutation of classical theory as a basis for his own theory (see quote above), and keyword data certainly confirm this. *Classical theory/doctrine/school* are frequent collocations, typically contrasted with the writer's own discourse and associated with refutation of their postulates. The words *school* and *doctrine* are almost exclusively used to represent classical theory and they are often accompanied by words like *accepted, dominant* and *orthodox*, only to emphasize the writer's *divergence* from it.

If we look at the concordance of *theory* throughout the book (246 occurrences, almost 20%), it is easy to see that *classical theory* is by far the most frequent collocation (53 occurrences, almost 20%), even more frequent than *the theory* (41). Postmodification with *of* is also confirmed to be very frequent (109), showing that *theory of* is followed by *employment* and *money* 13 times each, whereas *interest* is only present 3 times, but the emphasis lies rather on a theory of the *rate of interest* (19 occurrences) and falls equally on the analysis of a theory of *unemployment* (13). It is also interesting to notice that apart from *classical* (53), *general* (13) and *economic* (10), most other adjectives preceding *theory* are explicitly evaluative, and mostly of the kind that Hunston and Thompson (2000) would call evaluation in terms of "good" and "bad". Here is the full list:

bad, correct, faulty, foolish, nonsense, central complete, different (2), fundamental (3), foregoing (2), formal, independent, logical, ordinary, peculiar, preceding (2), prevailing, pure, scientific, separate, traditional (3).

Keynes' criticism of the classical theory is very explicit and centres on its oversimplification and faulty premisses: at the same time the presentation of his own model appeals to logical reasoning through formal refutation of the postulates of classical theory but also through his own forms of simplification. An adjective like *special*, for example, (key when compared with both present-day corpora) is repeatedly associated with *case* in criticism of the classical theory, but the second most frequent collocation is with *sense*, where it typically refers to Keynes' own *special definitions. Special senses are acceptable, but assumptions based on special cases are criticizable.*

Jucker 2007), typically meant to organize discourse, to establish and maintain relations between writer/speaker and reader/listener, or to manifest the value system of the speaker and the discourse community. In the context of argumentative discourse, keywords can be associated with culturally shared assumptions and values that constitute the implicit premises of argument within a socially situated argumentative practice (cf. also Rigotti & Rocci 2004). Both conceptual and organizational keywords may be a guide to the writer's evaluative position, and through this to the writer's position in disciplinary debates.

How much of this analysis depends on a close knowledge of the text is of course debatable. How much sense can one make of a keyword list without having a good familiarity with the text? Keywords, like most frequency data, point at elements that need to be explained, but part of the explanation is likely to be found in the co-text of the items, and ultimately in the text.

5. Overview of the chapters

The first section of the book explores the notion of keyness from different points of view. Michael Stubbs outlines the field from the point of view of language studies, discussing three loosely related uses of the term "keyword", as cultural keywords, as statistically meaningful repetition and as phraseological patterns involving extended units of meaning. His main theoretical focus lies on the critical link between words, texts and culture, while he argues for the need to relate words and texts to the social institutions which are characterized by texts and text-types.

The more specific problems and challenges of quantitative approaches, largely dominant in corpus linguistics, are presented by Mike Scott. The chapter maps out the problems of defining keyness, discussing statistical issues and the choice of a reference corpus, as well as illustrating issues of corpus stylistics.

One of the problems highlighted by Scott – the role of closed-class keywords – is picked up by Nick Groom and explored fully in a discourse perspective. The chapter presents the case for a specific focus on closed-class keywords as objects of corpus-driven discourse analysis. Their potential lies in the coverage they offer of phraseological data and in their capacity to reflect the constellations of meanings and values of a discourse community.

Jukka Tyrkkö draws a distinction between key words and keywords. Through the examples of hyperlinks in hypertexts, he claims that words may possess a degree of keyness due to their inherent markedness and their functional properties, rather than to statistical significance. Hyperlinks are paradoxically shown

The section closes with a chapter by François Rastier, who offers interesting reflexions on the background against which current research on keywords could be set by looking at the Web. He contrasts traditional programmes of knowledge representation with a corpus-linguistics web semantics – situating knowledge with texts rather than outside them – and advocates a re-thinking of the relationship between data and metadata.

Section II looks at keyness in specialised discourse. Martin Warren's text opens the section and links it to the first, by offering a new perspective on "aboutness". He looks at concgramming, identifying the most frequently co-occurring pairs of words, irrespective of constituency and/or positional variation. Analysis of the lexical concgrams looks at meaningful association to draw up a list of aboutgrams identifying the aboutness of a text. An examination of the text's phraseology and phraseological variation is shown to have great potential in defining the aboutness of a text.

The methodology is further discussed in Denise Milizia's analysis of the speeches of Tony Blair and George W. Bush. The focus of the study is first on the word *climate* and on the co-occurrence of *climate* and *change*. The analysis shows the importance of looking at phraseological units rather than individual words in looking for the aboutness of text.

Andrea Gerbig offers an interesting example of how different approaches can be combined in her study of a corpus of travel writing, from Early modern English literature to contemporary 'blooks'. Starting from statistically determined keywords, she studies key-keywords and associates, before moving on to contextual analysis of some words as extended lexical units and concluding with an analysis of keyphrases and phrase frames, thus including both repeated strings of words and repeated patterns.

Looking at phraseological combinations around selected keywords, Donatella Malavasi and Davide Mazzi study how different disciplines represent their own research activity, focusing in particular on subjects and objects of the activity, as well as on research procedures. By highlighting differences in the general lexis of self-representation in history and marketing, the study confirms the centrality of keywords in characterizing disciplines, as well as a considerable degree of inter-collocability between selected keywords.

Gill Philip looks at the problem of metaphorical keyness in a corpus of speeches by Italian female politicians. Starting from an identification of statistically generated keywords as mostly associated to a text's content, Philip looks for tools for the analysis of the relationship between keywords and the message of the text (covert keyness) focusing on evaluative language and metaphors. She sets

The third and final section looks at critical and educational perspectives. Erica Bassi studies how the *Kyoto Protocol* has been represented in two national newspapers: the Italian *La Repubblica* and the American *The New York Times*. Keywords are grouped into semantic fields to study the meanings associated to the protocol and closer analysis of words denoting 'disaster' and 'alarm' is carried out, emphasising the different strategies used by the two newspapers.

The study by Soon Hee Fraysse-Kim identifies keywords that trigger national consciousness of Koreans through an analysis of school textbooks used in elementary schools in four Korean communities: in South Korea, North Korea, Japan and China. The sense of homogeneity suggested across the politico-social borders is taken to reflect prevailing ideology, internalized and reproduced by school education.

Along similar lines, but moving towards pedagogical implications for literacy, Paola Leone uses keyness to identify the basic lexical patterns of school textbooks and matches them to the language young learners might be exposed to out of school. Results show discoursal, lexical, semantic, and morphological features which may be unfamiliar to the learner and should therefore deserve special attention in syllabus design.

The investigations presented in this book – originally presented at a conference held in Pontignano, Italy, under the title of the present volume – are quite narrowly focused on keyness in a corpus perspective, mostly involving attention to text and discourse. They are, however, illustrative of different topics, approaches, methods and theoretical assumptions. We are grateful to the contributors for this. Most of the contributions, on the other hand, have largely benefited from John Sinclair's ideas. We would like to add this volume to the long list of books dedicated to his memory, with gratitude.

References

Ädel, A. 2006. *Metadiscourse in L1 and L2 English* [Studies in Corpus Linguistics 24]. Amsterdam: John Benjamins.

Ädel, A. & Reppen, R. (eds.). 2008. *Corpora and Discourse. The Challenges of Different Settings* [Studies in Corpus Linguistics 31]. Amsterdam: John Benjamins.

Baker, P. 2004. Querying keywords. Questions of difference, frequency and sense in keyword analysis. *Journal of English Linguistics* 32(4): 346–359.

Baker, P. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.

Beaugrande, R. de & Dressler, W. 1981. *Introduction to Text Linguistics*. London: Longman.

Berkenkotter, C. & Huckin, T. 1995. *Genre Knowledge in Disciplinary Communication: Cogni-*

Biber, D., Conrad, S. & Cortes, V. 2004. If you look at: Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25(3): 371–405.

Bondi, M. 2006. A case in point: Signals of narrative development in business and economics. In *Academic Discourse Across Disciplines*, K. Hyland & M. Bondi (eds), 47–72. Bern: Peter Lang.

Carter, R. & McCarthy, M. 2006. *Cambridge Grammar of English*. Cambridge: CUP.

Cheng, W., Greaves, C. & Warren, M. 2006. From n-gram to skipgram to congram. *International Journal of Corpus Linguistics* 11(4): 411–433.

Conte, M. E., Petöfi, J. & Sözer, E. 1989. *Text and Discourse Connectedness* [Studies in Language Companion Series 16]. Amsterdam: John Benjamins.

Crismore, A. 1989. *Talking with Readers. Metadiscourse as Rhetorical Act*. New York NY: Peter Lang.

Dossena, M. & Jucker, H. (eds). 2007. *(R)evolutions in Evaluation*. Special issue of *Textus* 20(1).

Firth, J. R. 1935. Technique of semantics. *Transactions of the Philological Society*, 36–72.

Halliday, M. A. K. & Hasan, R. 1976. *Cohesion in English*. London: Longman.

Hoey, M. 1991. *Patterns of Lexis in Text*. Oxford: OUP.

Hunston, S. 2004. The corpus, language patterns, and lexicography. *Lexicographica* 20: 100–113.

Hunston, S. & Francis, G. 2000. *Pattern Grammar* [Studies in Corpus Linguistics 4]. Amsterdam: John Benjamins.

Hunston, S. & Thompson G. (eds). 2000. *Evaluation in Text*. Oxford: OUP.

Hunston, S. 2008. Starting with the small words: Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics* 13(3): 271–295.

Hyland, K. 2000. *Disciplinary Discourses*. Harlow: Longman.

Hyland, K. 2005. *Metadiscourse*. London: Continuum.

Hyland, K. & Bondi, M. (eds) 2006. *Academic Discourse Across Disciplines*. Bern: Peter Lang.

Martin, J. R. & White, P. R. 2005. *The Language of Evaluation: Appraisal in English*. Basingstoke: Palgrave Macmillan.

Moon, R. 2002. *Fixed Expressions and Idioms in English*. Oxford: OUP.

Phillips, M. 1989. *Lexical Structure of Text*. Birmingham: ELR, University of Birmingham.

Rayson P. 2008. From key-words to key semantic domains. *International Journal of Corpus Linguistics* 13(4): 519–549.

Rigotti, E. & Rocci, A. 2004. From argument analysis to cultural keywords (and back again). In *The Practice of Argumentation* [Controversies 2], F. H. van Eemeren & P. Houtlosser (eds), 903–908. Amsterdam: John Benjamins.

Schiffrin, D., Tannen, D. & Hamilton, H. (eds.). 2001. *The Handbook of Discourse Analysis*. Oxford: Blackwell.

Scott, M. 1997. PC analysis of key words – and key key words. *System* 25(1): 1–13.

Scott, M. 2008. *WordSmith Tools. Version 5*. Liverpool: Lexical Analysis Software.

Scott, M. & Tribble C. 2006. *Textual Patterns. Keywords and Corpus Analysis in Language Education* [Studies in Corpus Linguistics 22]. Amsterdam: John Benjamins.

Sinclair J. McH. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP.

Sinclair J. McH. 1996. The search for units of meaning. *Textus* 9(1): 75–106.

Sinclair, J. McH. 2004. *Trust the Text*. London: Routledge.

Sinclair, J. 2005. What's in a phrase. Lecture given at the University of Modena and Reggio Emilia, 15 November 2005.

Stubbs, M. 1996. *Text and Corpus Analysis*. Oxford: Blackwell.

Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Swales, J. 1990. *Genre Analysis: English for Academic and Research Settings*. Cambridge: CUP.

Vande Kopple, W. J. 1985. Some exploratory discourse on metadiscourse. *College Composition and Communication* 26: 82–93.

Wierzbicka, A. 1999. *Emotions Across Languages and Cultures: Diversity and Universals*. Cambridge: CUP.

Wierzbicka, A. 2006. *English: Meaning and Culture*. Oxford: OUP.

Williams, R. 1976/83. *Keywords: A Vocabulary of Culture and Society*. London: Fontana Press.

SECTION I

Exploring keyness