Knowledge organisation in LSP texts and dictionaries: A case study [*]

Natascia Leonardi
Dipartimento di Ricerca Linguistica, Letteraria e Filologica (DIPRI)
Universita' degli Studi
Macerata, Italy
leonardi@unimc.it

Keywords: LSP dictionaries, metalinguistic lexicon, conceptiual relations, terminology of linguistics, specialised knowledge

Abstract

In LSP dictionaries the specialised knowledge contained and organised in texts is selected and restructured. This paper is focused on the analysis of a case study: the *Dizionario generale plurilingue del Lessico Metalinguistico* (DLM – General Multilingual Dictionary of the Metalinguistic Lexicon). The dictionary of linguistics terminology under examination is planned to complement the reference products available in this area of knowledge. In fact, it has a particular outline as the materials it records are directly drawn from the most representative texts produced throughout the history of linguistic speculation. The plan of the DLM establishes that the terminological information stored (definitions, cross-references, formal variants, translations) is directly drawn from the original texts, and not elaborated by the compilers. Therefore, the definitions of the indexed terms are not produced by terminographers: they are 'defining quotations' identified and extracted by specialists from the source texts.

Specialised texts play an essential role in this project as they are analysed in order to both identify the core concepts used (or introduced) by their authors and to reconstruct the conceptual networks delineated in each of them. In the compilation of the DLM the problematic issues inherent in textual analysis clearly emerge. This is due to the fact that texts are multifaceted units where the various factors related to their structural organisation and informative content interact. The different degrees of 'density' of specialised information which is displayed in texts is determined, among others, by the conceptual, communicative, pragmatic, structural, cognitive, and socio-cultural components of LSP texts.

The procedures of retrieval and organisation of specialised knowledge carried out in the DLM project are analysed in this study through the consideration of a sub-section of its terminological inventory, i.e. the metalinguistic units extracted from a text in which focal linguistic issues are discussed. Although this book was produced in the pre-scientific period of the history of linguistics, it was chosen because, in addition to providing

---

[*] This article is based on a paper presented at the 2nd International Conference in the 360º Series "Encompassing Knowledge" that took place at the Århus School of Business, University of Århus, 15-17 May 2008.

interesting contributions to linguistics terminology – considered also from a historical viewpoint –, it yields a model for the arrangement of the conceptual relational network which is being implemented for the DLM. The bi-dimensional character of terminological records of the DLM is being integrated with graphic representations of conceptual relations, which provide a multidimensional outline to the defining section of this dictionary. The visual representation of relational networks provides further terminological information and it also makes available to the users an effective instrument for acquiring a more thorough understanding of the specialised knowledge which is transferred from LSP texts into this dictionary.

1 Introduction

The focus of this paper is on the LSP dictionary as a repository where knowledge drawn from specialised texts is organised. An essential aspect of this investigation is the way in which knowledge is distributed in the structure of the dictionary and can be accessed by its users. Indeed, the design and content of reference products depends on the purpose(s) they are intended to serve. The conceptual and linguistic information stored in specialised texts can be reorganised and presented in different ways when constructing LSP dictionaries, or terminologies.[1] Through the consideration of a case study this paper analyses the methods applied in the process of transposing information material from one text type to another, i.e. from the specialised text to the dictionary. In this framework the dictionary is interpreted as a type of text, of which different sub-types can be identified depending on their arrangement.[2] This implies that users can access information available in dictionaries according to different paths of knowledge retrieval and construction. The structural typology of specialised texts and dictionaries differs, hence a selection and reorganisation of conceptual and linguistic materials is necessary when the information contained in LSP texts is transferred into dictionaries.[3]

The specialised texts considered here belong to the field of linguistics and include essays and monographs produced by experts and intended primarily for experts or semi-experts. The present analysis focuses, in particular, on the way in which conceptual and terminological knowledge is dealt with in the blueprint and organising structure of the *Dizionario generale plurilingue del Lessico Metalinguistico* (DLM – General Multilingual Dictionary of the Metalinguistic Lexicon). The DLM is an ongoing project for the development of an on-line specialised dictionary of linguistics mainly addressed to specialists and advanced students (http://dlm.unipg.it/). The characteristics that distinguish this dictionary from other LSP reference works covering the same knowledge field will also be shown below through the analysis of a sub-section of the DLM (cf. § 4.1).

2 Features and structure of the DLM

---

[1] The terms 'LSP dictionary' and 'terminology' will be used as synonyms here; cf. Bergenholtz & Tarp (1995: 10-11, ff.) for a discussion on the relation between "specialised lexicography" and "terminology/ terminography". On this subject cf. also Brekke (2001).

[2] Dolezal (1989) and Wiegand (1990) highlight the textual dimension of general language dictionaries, but an analogous view can be drawn for those dealing with special languages. The specific organisation of an LSP dictionary is the outcome of different variables, such as intended users, purpose, dissemination medium, linguistic and domain knowledge provided or presupposed, etc. (Cf. Brekke 2001:181; Bergenholtz & Tarp 1995).

[3] The issues relating to the way in which the different aspects of specialised knowledge are represented in LSP reference works are analysed in depth in Schaeder & Bergenholtz (1994).

The elaboration of this LSP dictionary is based on a procedure of information retrieval from a corpus of specialised texts which differs from those generally used for dictionary production. In fact the DLM is intended to provide a lexicographic product entirely based on the terminological data available in the texts of the language sciences.[4] The starting point for term identification and analysis is the scrutiny of single works. Therefore, the overall corpus is not used by the compilers as a unique and comprehensive source of information, but is the final result of an assembly of distinct textual sources selected on the basis of their importance in the history of linguistics.[5] The final lexicographic product is a collection of the terminologies of outstanding authors, which are made available to the users of the DLM as integrated materials. In this way the outcome is a product in which the terminology of linguistics is represented both synchronically and diachronically, in a framework which takes account of the multifaceted reality of linguistics terms.

The compilers of the DLM are not terminographers but linguists, or occasionally students supervised by a specialist. In this project the interpretive role of the compilers is significant because it emerges through their direct intervention in different phases of the elaboration process. In the first place, compilers select the texts that will be analysed, then a textual analysis is carried out with the intent of identifying – through the application of the procedures explained below – the terms that are recognised as relevant for the theory discussed in that specific text. Terms and their corresponding original definitions are recorded in the DLM. In the working process the compiler extracts and analyses the terminology of one text at a time (including synonyms, variants, and translation equivalents). Definitions are pinpointed through a complete scrutiny of the work under examination and, particularly, of the passages where the selected terms are quoted (cf. § 3.1). The definitions recorded in the DLM are only those provided in a more or less explicit way by the author(s) who used or introduced the term. Hence they always consist of one or more quotations from the source text. As a result the defining sections contain quotations rather than definitions elaborated or re-elaborated by terminographers. This concise illustration of the compilation procedures of the DLM explains why, although it is called a *dictionary*, it should be interpreted as a *database of authors' terminology*.

Most terminological works are focused on a synchronic representation of the metalanguage of linguistics which is generally accepted by the community of specialists (e.g. cf. Crystal 1996). Whereas, the DLM was planned to make available to specialists and students a collection of the terminologies adopted throughout the history of linguistics by the most representative theorists. The decision to compile this type of dictionary can be explained by the need to develop a product that fills a gap in the panorama of the reference works available in this area of knowledge.[6] The peculiarity of the DLM consists in the fact that it combines the features of both a reference and a theoretical text, without displaying an encyclopaedic outline. A collection which provides the terms used or even created by single scholars gives a

---

[4] The model of the DLM was outlined by Cristina Vallini with the aim of providing a systematic gathering of the terminological heritage of the language sciences. The principal characterising feature of this lexicographic resource is that it is grounded on strong theoretical bases provided by the special typology of the materials that constitute it (cf. De Meo & Lorenzi 2006:§1). A comprehensive description of the DLM is given in Lorenzi (2002).

[5] The structure of the corpus cannot be described in full, but a basic outline is provided below (cf. §3).

[6] For example, as regards contemporary English-language dictionaries of linguistics, Crystal (1997:26) highlights the absence of research which could provide a "list of historical sources or corpus citations". On the other hand, a corpus of the principal linguistics texts is the aim of the project CTLF (*Corpus de Textes Linguistiques Fondamentaux*: http://ctlf.ens-lsh.fr/), but for the moment this resource does not have a lexicographic outline.

representation of the spectrum of the focal concepts of individual theories of language. In addition, it makes available information on the fundamental issues characterising the different periods in the history of linguistic thought.

The diachronic dimension – which is given prominence in the DLM – may play an important role in terminology, especially in the humanities.[7] Linguistic concepts and terms outline conceptual structures, i.e. theoretical networks, which are specific for authors and schools. Therefore, a terminology of linguistics needs to cope with variations and changes displayed by terms when used by different authors and/or in different periods of time. In the traditional lexicography of linguistics, the synchronic perspective is dominant. Nevertheless, it is often integrated with encyclopaedic notes which provide the end user with references to specific analytical frameworks and/or author's usage. These are features which may help in the definition of particular values acquired by terms within specific theories. The outline of the DLM favours the handling of terminological complexity, thus the ambiguity, variation, and redundancy that often characterise linguistics terms can be clarified.[8]

3 The distribution of conceptual knowledge in texts

In the following passages I will focus my attention on the importance of original texts in the production of the DLM and on the problem of the interpretation of concepts which has a particular character in this dictionary because it is basically related to the selection (and 'presentation') rather than to the 're-elaboration' of defining materials. A specialised dictionary is compiled in order to provide a representation of expert knowledge, and the corpus-driven dictionary under examination here makes available first-hand specialised knowledge in a direct way. The terms recorded in the DLM are identified in texts produced by outstanding authors throughout the history of linguistic speculation. Hence it also includes the works written in the 'pre-scientific period'; in fact, my attention will be focused on one of these. The DLM also contains data from dictionaries and encyclopaedias of linguistics, but limits them to the entries while leaving the definitions aside. This type of source is mainly used for making available information on multi-language terminological equivalences but cannot offer original data on concept definitions as these are produced by lexicographers (cf. De Meo 2002:43). Complete data on inter-linguistic concept equivalents are included in the entries of the DLM only when they can be drawn from official translations of the original works or from translations occasionally provided by the author within the text under analysis. Antia – among others – underlines the close connection between designations and the conceptual knowledge they convey:

> As labels (linguistic and non-linguistic) for specialised concepts, terms are a means of acquiring, retrieving, creating, communicating, storing, representing and operationalising specialised knowledge (Antia 2000:xv).[9]

---

[7] Nevertheless, the prominence of the diachronic dimension of terms also in hard sciences is underlined in terminology literature (cf. e.g. Temmerman 2000:107-115).

[8] Cf. Gotti (2003:46-65) for an analysis of the characteristics of specialised discourse which challenge the basic principles of terminology though being part and parcel of the actual nature of specialised lexica.

[9] Numerous definitions of the concept of 'term' have been given in the literature; these reflect the complexity of this notion and its various interpretations in different theories. A basic distinction can be made between those readings which interpret the term merely as a designation (e.g. a "lexical unit consisting of one or more than one word which represents a concept inside a domain", cf. *Glossary of terms used in terminology*, s.v. "term") or as a more complex entity which embodies the referential, linguistic, and conceptual facets of specialised knowledge (e.g. Cabré defines terms as "three-way units of meaning (thing-name-meaning) which refer to the specialized reality", 1996:19). Here 'term' is used with both meanings but, when it is necessary, distinguishing designations are adopted instead.

In the DLM the compilers identify as 'terms' all the lexemes which are attributed a metalinguistic value by the authors of the texts considered. Thus, it also records terms that are often excluded from traditional specialised dictionaries, such as similes and metaphors.[10] Therefore, in addition to the terminological and conceptual data which are usually represented in LSP dictionaries, the DLM also takes account of those that belong more directly to the communicative facets of knowledge stored in specialised texts, i.e. information which normally does not appear in the conventional outline of terminologies.

The specialised contents structured in texts are necessarily synthesised and rearranged in terminologies. This process serves to provide the end user with a satisfactory representation of specialised knowledge which makes it easily accessible and usable for specific purposes (e.g. related to the reception, production or translation of texts). In fact, the texture of specialised concepts displays a relative complexity within texts, where the units of knowledge are distributed and variously organised. Different degrees of informative density can be recognised throughout the textual passages. Jansen (2003) establishes three main parameters for detecting the density of information in LGP texts; but these can be applied to the analysis of LSP texts as well. The distinguishing criteria consist in the quantity of information, the quality of information, and the quantity of non-information contained in texts (Jansen 2003: 9-11; 212-213). The first parameter refers to the way in which content data are expressed in texts, that is to say it considers whether information in the surface structure is implicit, explicit or omitted. The second principle concerns the degree of compactness characterising the linguistic expression of relevant information throughout the text, i.e. the degree to which linguistic materials express informative contents in a distributed or concentrated way.[11] The third parameter pertains to the quantity of 'non-information' present in texts, namely linguistic materials which do not perform an informative function but rather discursive, structuring, evaluative, and interactional ones (Jansen 2003:9-11; 212-213).

Jansen applies these criteria to the analysis of narrative texts, while in this study the 'density of information' refers to linguistics texts. Hence as regards these principles, and particularly the third one, it needs to be highlighted that the importance of the functions of linguistic and non linguistic materials depends on the typology of texts. Since the present scrutiny is devoted to the organisation of specialised knowledge in terminologies, and specifically in the DLM, the identification and extraction of terminological and defining materials in LSP texts are focal issues. The analysis of texts is driven by the intent of identifying core concepts and reconstructing conceptual patterns, i.e. threads and networks of concepts. Thus the content (or conceptual) dimension of texts acquires a crucial role in this process, but a consideration of the cognitive, functional, and communicative components that contribute to shape LSP texts is also essential.

An analysis of LSP texts can be characterised by different but complementary methodological perspectives which give account of the complexity of specialised communication, so much so that Baumann (1992) introduced an integrated approach to the study of specialised text

---

[10] These points are discussed by De Meo (2002:35; 49-50) and De Meo & Lorenzi (2006:§3). As an example can be quoted the similes of the "jeu d'échecs" and the "feuille de papier" introduced by Saussure (1916:43; 157) when discussing the concept of *langue*. In the DLM a metalinguistic value is attributed to these similes, which are included as entries and connected to "*langue*".

[11] Jansen refers in particular to lexical and syntactic compactness, but for the purpose of the present study this criterion can be extended also to longer textual sections (i.e. paragraphs, chapters, etc.).

linguistics. His theory considers the various factors that combine to form the specialised knowledge contained in texts.[12] As Rothkegel synthesises (2005:84):

> A text is a communicative unit that converts general schematic knowledge into usable information in a sense that persons are involved into some communicative processes in which the application of knowledge plays a crucial role and which includes cognitive states, attitudes, goals, emotions. This is a new quality of 'information processing' because it is embedded into the physical, mental, social, and institutional life of human beings. As a consequence, a text is considered to be an organisational device for communicating knowledge according to some specified functions which are related to some specified situations of communication.

The articulation of specialised communication is also a central issue in terminology studies, as the recent currents of research attest (cf. e.g. Cabré 1999, Temmerman 2000). The standpoint of text analysis directed at terminological investigation takes into consideration the different communicative aspects that interact in specialised texts:

> A text is not just a linguistic unit but a mode of social and cultural expression allowing individuals to relate to one another. Therefore, many aspects besides those that are purely linguistic must be analyzed in order to characterize a text correctly. First, texts are complex linguistic units and conform to the rules of combination of each language system. Secondly, texts are complex pragmatic units, because they are produced by people who are neither psychologically transparent nor ideologically neutral. Thirdly, texts are complex sociolinguistic units because a language is a system for social communication that occupies a certain place in society that uses it and has a relationship to other languages and their societies with which it is in contact. Finally, texts are complex cultural and anthropological units which reflect and communicate a system of cultural and ideological values by means of discourse (Cabré 1999:57).

In the blueprint of the DLM, the communicative issues that coalesce in texts, and that contribute to outlining its knowledge architecture, are interpreted as effective components of specialised communication. For this reason they are seen as potential sources of relevant metalinguistic information. Nevertheless, in the textual passages where the metalinguistic function is not explicitly activated, specialised information usually appears as 'diluted' and its threads need to be drawn together by the compiler of the DLM who aims to select and record terminological data.

A complete scrutiny of this subject goes beyond the scope of this paper; the aim of this section is limited to highlighting the interpretive problems encountered by the receivers of texts and, in particular, the compilers of the DLM. The latter have to identify key concepts (and possibly their various designations) and their clear-cut definitions in linguistics texts. Hence their work is characterised by the particular attention they have to pay to the

---

[12] Baumann (1992) proposes an integration of the methods of different disciplinary fields (e.g. sociology of language, psychology, terminology, etc.) as these may explain the facets of specialised communication that can be identified at different levels through a structural and functional analysis of specialised texts. Cf. also Baumann (2001a, 2001b) for a discussion of the aspects which contribute to shaping knowledge systems in specialised texts, and which determine the necessary application of a combined perspective in textual analysis.

conceptual information structured in texts, but in order to do this, they also have to consider stylistic, communicative, and interactional components of the linguistic message.

3.1 Definitions in linguistics texts and in the DLM

The DLM was planned as a particular type of reference work: since its initial outline it has not been conceived as a dictionary of linguistics that would contain the product of the elaboration of specialised materials made by terminographers. It is rather intended as a repository of selected original data organised in a format which can make them easily accessible for students and specialists (De Meo 2002:35-38).

The various types of specialised reference works – ranging from the glossary, to the dictionary, the terminological database, and the terminological ontology – provide underlying structured representations of the conceptual relational system designated by the terminology of a specific discipline or field of knowledge. Different strategies are used for reorganising and presenting knowledge in the various models of specialised reference works. In any case, the identification of the conceptual system is a necessary working step even when it is not directly made available to the user.

As discussed above, the specialised knowledge that can be found in texts is often 'distributed' in discourse. The works which make up the corpus for the terminological research of the DLM mainly consist of monographs and, to a lesser extent, of articles; hence the corpus is basically made up of informative or argumentative texts. The communicative purpose, and the structural organisation of texts tend to enhance the fragmentation of relevant information in different prose passages. In our perspective this implies that the defining properties of a concept are often distributed in various parts of a text. In addition, a text – or particular passages of it – can be characterised by different degrees of 'density' of explicit specialised contents, depending on its typology and the readers it is addressed to.

In linguistic writings the use of terms made by specialists is not always accompanied by a gloss or an explicit definition; often long passages – or even whole chapters – are devoted to the discussion and the thorough analysis of a concept. Hence, in the flow of discourse conceptual information is provided, but the reader has to 'reconstruct' a concept definition by gathering and organising partial information distributed in different sections of the text.

The extraction of a single sentence or paragraph containing a complete and fully satisfactory definition is a difficult – sometimes impossible – task for the compilers of the DLM. For this reason, in many cases more than one quotation is provided for the same term. In this way the information relevant for identifying the concept is selected and extracted from the source text and made available to expert or semi-expert users. Hence the concept can be 'reconstructed' from the information given in the textual passages, with the 'side benefit' of also retrieving data on the usage of terms in context. Occasionally the dilution of information makes it impossible for the compiler to identify an informative quotation. In this case, the guidelines of the DLM establish that the term cannot be recorded, even though the concept might eventually be delineated through a close scrutiny of the text.

Quite rarely those collected in the DLM have the characteristics of satisfactory definitions, that is to say analytical ones, or anyway definitions that provide the necessary and sufficient properties of the concepts in question, as those obtained for example through the application

of theoretical perspectives different from the logical and ontological ones. In fact, the type of definition depends also on the nature of the concepts that are defined.[13] An alternative approach is represented, for example, by conceptual analyses based on prototype structure, cognitive models, or a diachronic scrutiny.[14] The structural and communicative features of linguistics texts and the specific nature of at least a part of the concepts introduced there may account for the definitions which do not match the analytical model. The following definition of "meaning" in Simon Dik's *The theory of functional grammar* (1989) could be considered an example of a rather satisfactory term definition given in the DLM:

> […] I shall reserve the term "meaning" (or "semantic content") for the information which is in some way or other coded in the linguistic expression as such. Thus, semantic content is a feature determined by the language system, to be accounted for in the grammar of a language"
> (Dik, 1989:11; http://dlm.unipg.it/Consultazione/citazioni.asp?IDCitazione=14615).

Here the author also provides a clear synonymic correspondence – recorded in the dictionary – between "meaning" and "semantic content".

But in most cases, those provided in the DLM should be identified as 'defining comments' rather than proper definitions (cf. § 4.1). For example, the concept "meaning" as intended by Bloomfield in *Language* (1935) is illustrated by seven quotes which concurrently contribute to explaining the term under examination by providing its characterising features, but none of them if singled out can be considered a satisfactory definition. The DLM was given the particular outline shown above because it was designed to provide a valid source of metalinguistic information which is faithful to the original speculation of the linguists whose works are considered. And above all, it is proposed as a reference work that can take account of the multifarious reality of authors' usage of terms, also from the historical perspective.

Since the materials of the DLM make available first-hand knowledge drawn from specialised texts, its users can compare the different values attributed to the same term by various authors in different periods – that is to say, the use of the same designation for distinct concepts in one knowledge domain – or also the evolution of and change in concepts and designations over time. For example, in the text by Wilkins that I have analysed (cf. § 4.1), we find the term "semivowel" (Wilkins 1668:360), though it refers to a concept which does not correspond to that which we associate today with the same designation. Wilkins introduces this term when he signals that his contemporaries use it for referring to "breathed consonants".[15] Nevertheless, he also discusses the concept equivalent to the modern content of the designation 'semivowel', but the designation he uses is "Intermediate Vowels", or its

---

[13] Temmerman introduces the notion of "units of understanding", which overlap only in part with the traditional interpretation of the 'concept', as these include also 'categories' (Temmerman 2000:73; cf. also *ibid.*:43, 65). Concepts can be defined according to logical and ontological analysis, while 'categories' have a prototypical structure and the theoretical parameters that are suitable for their definition are different from those used to define concepts.

[14] Geentjens *et al.* (2006:10) underline that terminological analysis in the framework of the "sociocognitive approach" is based on the methods used in cognitive semantics, i.e. "prototype structure analysis, cognitive model analysis and diachronic analysis" as these theories can explain how human knowledge is acquired and organised.

[15] "SEMIVOWEL: All simple letters may be distinguished into such as are [...] Intercepted and shut according to degrees Lesser; which because they have something Vowelish [sic] in them, are therefore by some styled Semivowels, being spirituous and breathed [...]" (Wilkins 1668:360). These sounds would be designated in contemporary phonetics terminology as nasals, fricatives, retroflexes, and laterals.

synonyms "[Sounds having a] Middle nature / mediæ potestatis / middle power" (Wilkins 1668:360, 370).

This example gives an idea of the changes which characterise the terminology of a discipline from a historical perspective, and the DLM with its focus on the terminology of single authors can give account of such data. Linguistics is one of the disciplines in which the socio-cultural and historical dimensions play an important role in term formation. As shown in the example above, secondary term formation often consists in a reformulation of the concept which is made to correspond to an unvaried designation due to changes in the background theory. So, for instance, 'sign', which is obviously a basic notion in linguistics, corresponds to different concepts in this and in interrelated fields – also from a synchronic perspective – depending on the theoretical approach adopted (e.g. Saussure's 'sign' is not the same entity as that intended by Ogden & Richards, 1923).

4 The structural organisation of the DLM: consultation paths and conceptual relations
The DLM is a database which can be consulted following different paths as the search can be by lemma, author or text. The basic organising principle followed in its elaboration is the indexing of terms starting from the specialised text in which they are used. The search by text gives access to the comprehensive alphabetical list of the terms extracted from it; from the list the specific record where each of them is analysed can be consulted. If the search is by lemma, the result is a list of its occurrences in all the indexed works. From this overview users can consult the single records of the occurrences, and compare the definitions or the uses which characterise the various authors or, if it is the case, their uses in different texts by the same author. The search by author provides the corresponding indexed text(s), which can be accessed with the system illustrated above.

These types of searches are integrated by the possibility of retrieving data organised according to a chronological principle. This search can be by year or lemma. In the first case, the result is an overview of the terms recorded in the DLM according to the year of their appearance in specialised texts. The chronological search by lemma yields a scheme regarding a single term showing the year(s) of its occurrence in the indexed works; from the general graphic it is also possible to access each single record.

This overview of the DLM points out the peculiarities which distinguish it from the repertoire of electronic or printed dictionaries and encyclopaedias of linguistics which it sets out to complement. In fact, as underlined above, its principal aim is not so much that of satisfying a need for clear-cut definitions or translation equivalents, as these are provided by other types of reference works. Rather, it is focused on making available an organised repository of original and comparable data.

At this stage the DLM is not complete, hence we cannot consider its corpus as a satisfactory spectrum of linguistics knowledge. Indeed, my attention here is not focused on this aspect but on the potentialities that its structure displays, especially as regards the type of knowledge represented in this terminology. The DLM proves that the nature of the specialised knowledge provided by a reference work also depends on the materials selected and the way in which they can be accessed by the users. In fact, the type of knowledge contained in a terminological tool and the format of its presentation can affect the acquisition and structuring of specialised information in users' minds.

The functions of the DLM now available to the public do not yet include the possibility for end users to fully access terminological relations, apart from synonymy, generic cross-references, and translation equivalences. Since in this dictionary the single text is the starting point for terminological analysis the relations now displayed are established among the terms of each single text. The same principle will also apply to the future implementations aimed at providing a more complete overview of the conceptual relations existing among such terms. These will soon be available to the public on the DLM web-site.

The changes and new applications under development are directed at integrating the indexed materials with relational schemes which highlight the connections among the terms recorded for each text. As will be exemplified below, a relational conceptual scheme will complement the information that can be drawn from the 'defining quotations' in each entry. With this new component the bi-dimensional model of the DLM that provides terms and authors' definitions is projected in a multidimensional scheme where the user can retrieve more complete information concerning the indexed terms. The relations identified and outlined in the visual representation include those that can be directly drawn from the 'defining quotations' of the term under analysis, and those that do not appear in an explicit way in the quotations but can be elicited from the definitions of related terms. If it is necessary, the whole text becomes a source of further knowledge that can support the compiler in this interpretive task.

The graphic scheme now planned is probably not the best possible one as it displays an "associative system" (cf. Wright 1997:90). One drawback of this type of representation consists in the fact that all types of term relations – indicated by connection lines in the graph – are put on the same visual level. Still, this flaw is balanced by the presence of descriptive labels which make explicit the types of connection indicated by the lines, and different colours are used for each of them. In addition, the presence of directed arrows where appropriate in the graph also shows whether the relations are symmetric or asymmetric (cf. Figures 1-3 below). The relations that can be displayed on the same graph are both hierarchical (hyperonym-hyponym, part-whole) and non-hierarchical ones (opposition, derivation, synonymy, translational equivalence, etc.).

A comprehensive conceptual network for all the terms in the DLM has not been planned, as its use would be impractical and theoretically inappropriate. This is mainly due to the fact that the leading classificatory principle of the DLM is the analysis of the set of terms used by single authors over an extremely long period of time. Furthermore, an overarching graphic representation could not easily cope with the polysemy and variation that terms display, especially when they are analysed from a diachronic perspective. Hence for each text a graph is delineated as a result of the compiler's identification of specific conceptual relations. This phase of the work is an additional component of the compilation process illustrated above (cf. §2). The outline of a conceptual network is made possible by the close scrutiny of information gathered through the previous steps of textual and terminological analysis. Thus the relational graph is not limited to offering a schematic view of the data contained in the corresponding terminological record: it can also provide additional specialised knowledge. In this way it becomes, in all respects, an effective component of the defining section of the DLM.

Users can access the graph through a link that will be provided in each record; the result of his search will give a partial view of the complete chart, i.e. the section centred on the search term showing all the concepts it is directly related to. In fact, a complete view of the graph representing the entire terminology of a single text would be difficult to consult, especially when the number of terms indexed is very high. Yet, from the partial graph the user can

continue the exploration: by clicking on the terms related to the search one the graphs devoted to each of them can be consulted.

4.1 The relational scheme of the *Essay* in the DLM
Through the consideration of text which I have analysed and indexed for the DLM, it is possible to illustrate more clearly two focal features related to knowledge retrieval and representation in this LSP dictionary. The first point concerns the 'density' of conceptual information contained in textual data and, particularly, the difference between the analytical definitions of terms directly given in LSP texts – which are relatively rare – and what might be called 'prose' (or 'discursive') 'definitions'. The latter are discursive sections or even chunks of text containing relevant conceptual data which are therefore not available as a cohesive whole in a single passage, but rather dispersed throughout longer textual excerpts. The second issue relates to the role that the abovementioned data play, on the one hand, in making available the conceptual information conveyed by terms and, on the other hand, in the function they have in the visual representation of such data through a graphic relational system.

The text under examination here is *An Essay Towards a Real Character and a Philosophical Language*, published by John Wilkins in 1668 under the auspices of the Royal Society. Even if this work belongs to the pre-scientific period of the history of linguistic thought, it provides interesting data which are relevant both for the content and the organising principles of the DLM. Indeed, Wilkins' text represents a contribution to the specialised language and concepts of linguistics of the end of the 17th century (cf. Leonardi 2008). In addition, it displays remarkable features with regard to the principles of conceptual classification and definition; this is a basic issue in terminology science, and is also particularly important for the delineation of a relational conceptual model capable of organising the DLM terminology.

The *Essay* can be considered 'anomalous' if compared with the other texts indexed in the DLM. It is not possible here to go into the details of the structure of the *Essay*, but it is necessary to point out that its textual architecture is particularly complex as it includes different – and partly cross-referenced – types of texts. Each of them provides a distinct type of definition which have now all become 'defining quotations' in the DLM. '*Discursive definitions*' can be found in the prose sections devoted to Wilkins' consideration of various aspects of language which are significant in the history of linguistic thought. This section does not explicitly aim at providing concept definitions but is anyway extremely important for the DLM terminological analysis. In addition, this work contains an *Alphabetical Dictionary* where *glosses* are used for disambiguating polysemous words or homonyms. Yet, the primary defining section of Wilkins' dictionary is arranged according to 'philosophical' principles, that is to say following a model which aims at reproducing the organisation of human knowledge. In this section *analytical definitions* are given because this part of the text is structured as an ontology, a hierarchical and network scheme, based on three main classificatory levels which display the properties of the "things and notions" categorised in Wilkins' book. Hence the conceptual information contained in the defining section of the *Essay* is provided by an interplay of devices and types of text: it integrates the traditional structure of an alphabetical dictionary, a clear-cut ontological frame, and prose discussions containing relevant conceptual data.

From the overall inventory of concepts defined in the *Essay* I have extracted and analysed the terms related to linguistic issues. Some parts of the relational graph I elaborated for

structuring Wilkinsian terms in the DLM are basically drawn from the author's own scheme; as explained below, also part of the terminology used for specifying the relations is borrowed from that used in the *Essay*. I integrated the relations explicitly established by the author with others that can be clearly identified in the text. An interesting point of this contribution to the DLM is that, in this case, the originality of the materials is not limited to the 'defining quotations' but also involves the relational graph, whose delineation otherwise is left entirely to the interpretation of the compiler. In fact, the terminology of the *Essay* is also used as test data for the implementation of the new cataloguing and consultation functions of the DLM: the peculiar textual typology of this work proposed, at the end of the 17th century, a relational defining structure similar to that planned for this contemporary terminology.[16]

The following examples show the importance of a relational scheme that integrates the conceptual data expressed in different sections of a text, the *Essay*, which in itself displays distinct types of textual structure and of density of information. In its relational graph coalesce conceptual data retrieved from explicit (and deliberate) term definitions, brief glosses, and 'prose definitions'. The term "language" is defined in the *Essay* as follows:

> Of DISCOURSE, Or the several notions belonging to *Grammar* or *Logick* […]. To which [i.e. Discourse] may be annexed that particular way of discourse, most in use, namely by articulate voice and words, called LANGUAGE. Tongue, speech, linguist, dialect (Wilkins 1668:44).

An outline of the conceptual relations of the term "language" which is based on the analytical definition given in the ontology provides only a limited view of this concept as it is intended by the author. In its corresponding graph only four types of relations would emerge (*is-part-of*, *has-adjoined*, *term-connected-with*, *is-affinis-to*), these connect the term "language" with eight concepts (discourse, grammar, articulate voice, tongue, speech, linguist, dialect), as Figure 1 shows.[17]

If the information on "language" presented above is integrated with the data resulting from the analysis of other terms in the text, additional relations can be identified, and the conceptual network becomes more complete: six relations connect twelve concepts (cf. Figure 2). Synonymy and hyponymy (*is-synonym-of*, *is-a*) are added to the relations identified in the initial definition. In fact, in the complete graph of the term "language" its synonym (tong) and hyponyms (mother tongue1, mother tongue2, dialect1, derivation1) also appear.[18]

The denomination of some relations might seem peculiar, my reference is in particular to "is-affinis-to" and "has-adjoined". This is due to the fact that five basic relations are indicated by Wilkins himself in his defining ontology. For the DLM the same designations used in the *Essay* – with slight reformulations – were maintained, as these can be basically understood also by users who are not familiar with Wilkins' work.

---

[16] On the defining architecture of the *Essay* cf. Hüllen (1999, 2004:284-287, 290-291); Maat (2004:135-217).

[17] In particular, "language" *is-part-of* "grammar", *is-affinis-to* "discourse", is a *term-connected-with* "articulate voice", and *has-adjoined* terms, i.e. "tongue", "speech", "linguist", "dialect".

[18] Homonymic term designations are distinguished by a number; formal variants are considered as complete synonyms.
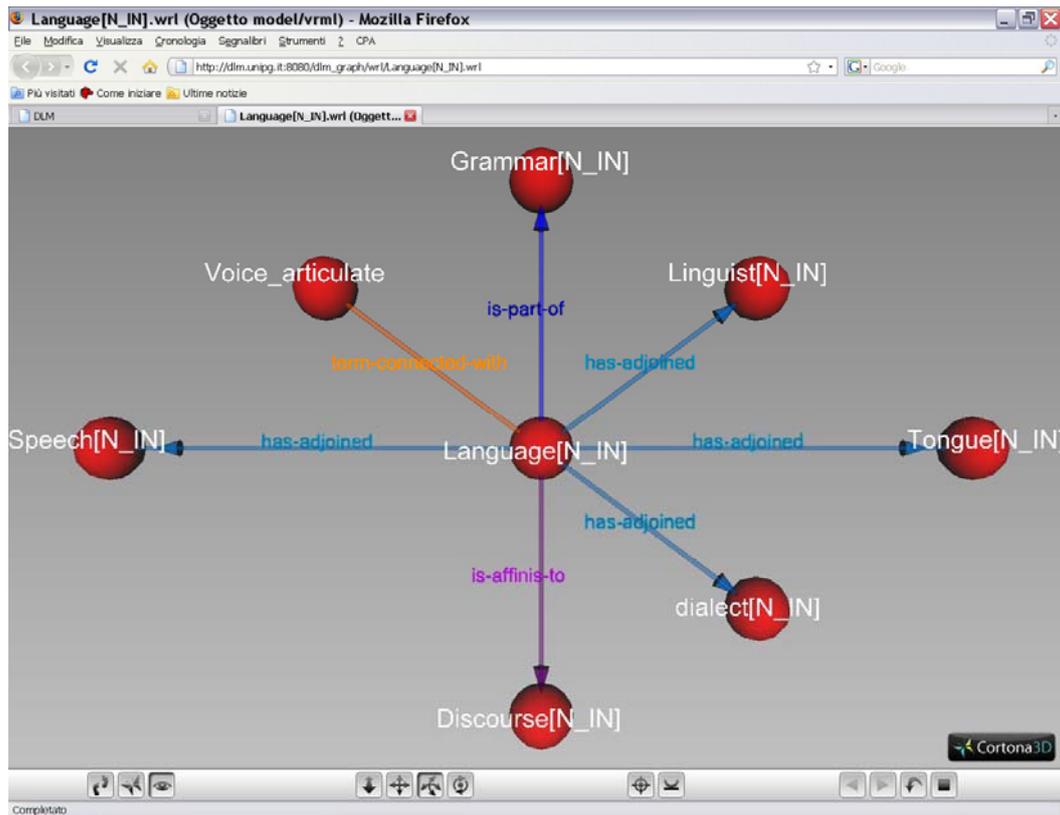
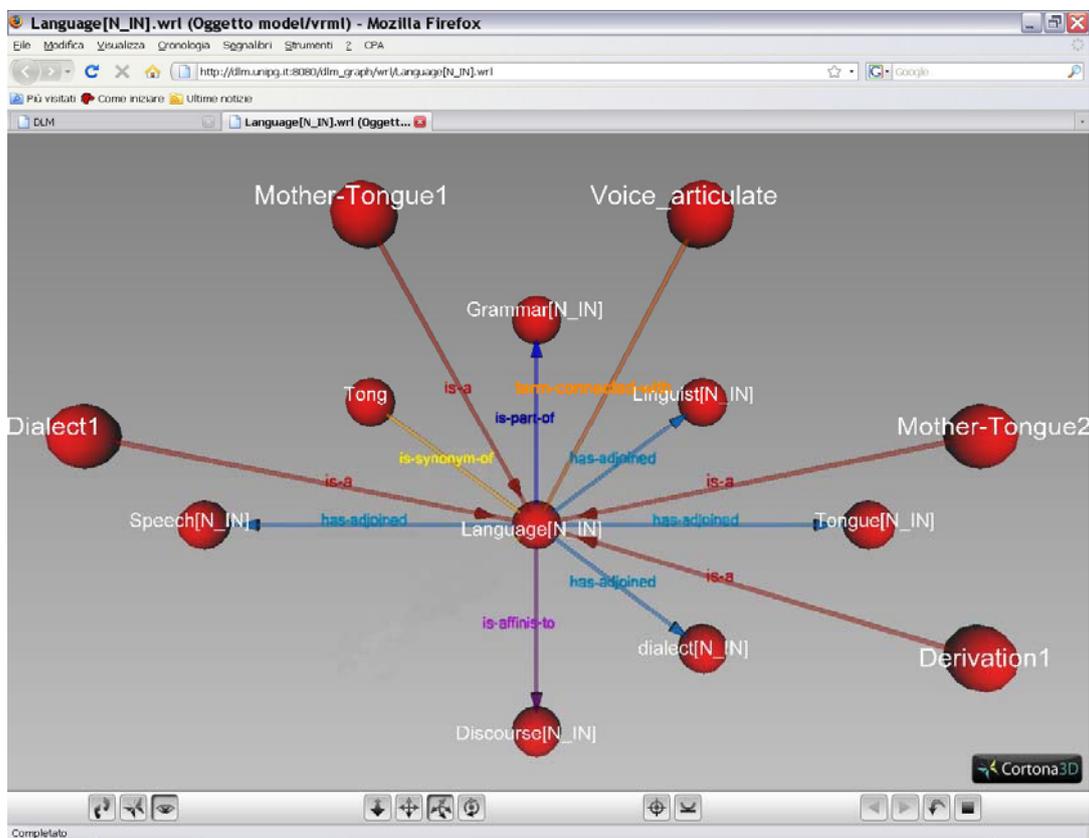**Figure 1:** Graph of the basic relations of the term "language".

Figure 2: Graph of the complete relations of the term "language".

The relations identified by Wilkins are: *is-a* (alternating with *is-a-type-of*) and *is-part-of*, and the rather general ones of opposition (*has-opposite*) and affinity (*is-affinis-to*): *has-opposite* generalises in the *Essay* and in the DLM various types of opposition (complemenaries, antonyms, reversives, converses, etc.) and *is-affinis-to* indicates, in a general way, a privileged relation with a co-hyponymic concept.[19] A further type of association indicated by the author is called *adjoined* (by focusing on the core term I renamed it *has-adjoined*), this tag includes both complete synonyms and more often hyponyms of the head-term. An example of 'adjoined' terms is given in the definition of "Language" where "tongue, speech, linguist, dialect" are connected to the main term. The specific relation between a term and the 'adjoined' ones is usually made explicit in the *Alphabetical Dictionary* attached to the *Essay*. The relations established by Wilkins were not sufficient for describing the connections between concepts in the network, for this reason I added nine more: *is-synonym-of* (which indicates complete synonymy or variation of graphic form), *has-property* (which serves to detect distinctive features, especially those of co-taxonyms), *derives-from*, *translation-of*, and *term-connected-with* (which signals a generic conceptual relation and also a cross-reference to other terms). In addition, the following relations have been added, even though they apply to a rather limited number of term relations: *is-cause-of*, *group-of*, *location*, and *people* (these are mainly used for the description of terms designating natural languages).

The following is an example of how the graphic representation of conceptual relations requires a knowledge which is often not made explicit in the 'defining comments' provided in specialised texts. Hence a graph can also occasionally show data which are not included in the 'defining quotations' of that specific term, either because this information comes from the definitions of other terms or because they are included in the original text, but are available only in sections or passages which do not match the typology of a 'defining quotation'.

The interpretation process that leads to the compilation of the record and the graph of the term "Synonym" (and its variant "Synonymous") takes into consideration different sections of the *Essay*. It is not defined in Wilkins' ontology but three 'defining quotations' were detected in the prose text and in the *Alphabetical Dictionary*:

> In respect of *Synonymous* words, which make Language tedious, and are generally *superfluities*, since the end and use of Speech is for humane [sic] utility and mutual converse; *magìs* [sic] *igitur refert ut brevis, & rectus, & simplex sit quàm* [sic] *lòngus* [sic] *& varius* […]. And yet there is no particular Language but what is very obnoxious in this kind" (Wilkins 1668:18).

> "Many of the Synonymous words put to the Radixes [in the *Philosophical Tables*], are referred to more heads then one, upon account of their various equivocal acceptions. And besides such words or phrases as are more plainly Synonymous […]" (Wilkins 1668:290).

> "Synonym [Of same meaning]" (*Alphabetical Dictionary*, s.p.).

These passages only reveal one relation of synonymy and three of the general type '*term*[s]-*connected-with*': such data are not adequate to provide a satisfactory definition. Yet, in the

---

[19] The formula commonly used by the author is '[term] x to which [term] y is affinis' or, more rarely 'is annexed', as in the citation above (Wilkins 1668: 44). A similar formula is used for the relation of opposition: '[term] x to which [term] y is/may be opposed'. Often formulas are not used at all, and the information is provided in the corresponding term entry of the *Alphabetical Dictionary*.

relational graph one distinguishing property of this term also emerges and, what is more relevant, it is identified as a "Defect of words" (cf. Figure 3). The recognition of its hyperonym is crucial for the definition of this term in Wilkins' linguistic theory. Such information, which does not appear in the 'defining quotations' recorded in the DLM, can be easily retrieved by the compiler. The problem is that the textual structure which contains it does not match the requirement of the DLM guidelines concerning 'defining quotations'. Indeed, it is contained in a passage which introduces the discussion on what are identified as "linguistic defects", which reads: "Besides these defects in the usual *Alphabets* or *Letters*, there are several others likewise in the *Words* of language, and their Accidents and Constructions" (Wilkins 1668:17). The second point in a list of defects is one of the passages cited above. Moreover, the discussion that Wilkins makes on two pages (Wilkins 1668:17-18) contribute to clarifying the nature of this concept as one of the defects of words. This term is also explicitly related to "meaning" in the gloss of the *Alphabetical Dictionary*. On the other hand, "synonym" is not mentioned at all in the definition of "meaning" provided in the ontology of the *Essay*.[20] Hence the cross-reference between these terms is made easily available to the users of the DLM by the implementation of a system of graphic representation. In fact, this relation clearly appears in the graphs of both terms, but it is evident only in the defining record of "synonym".
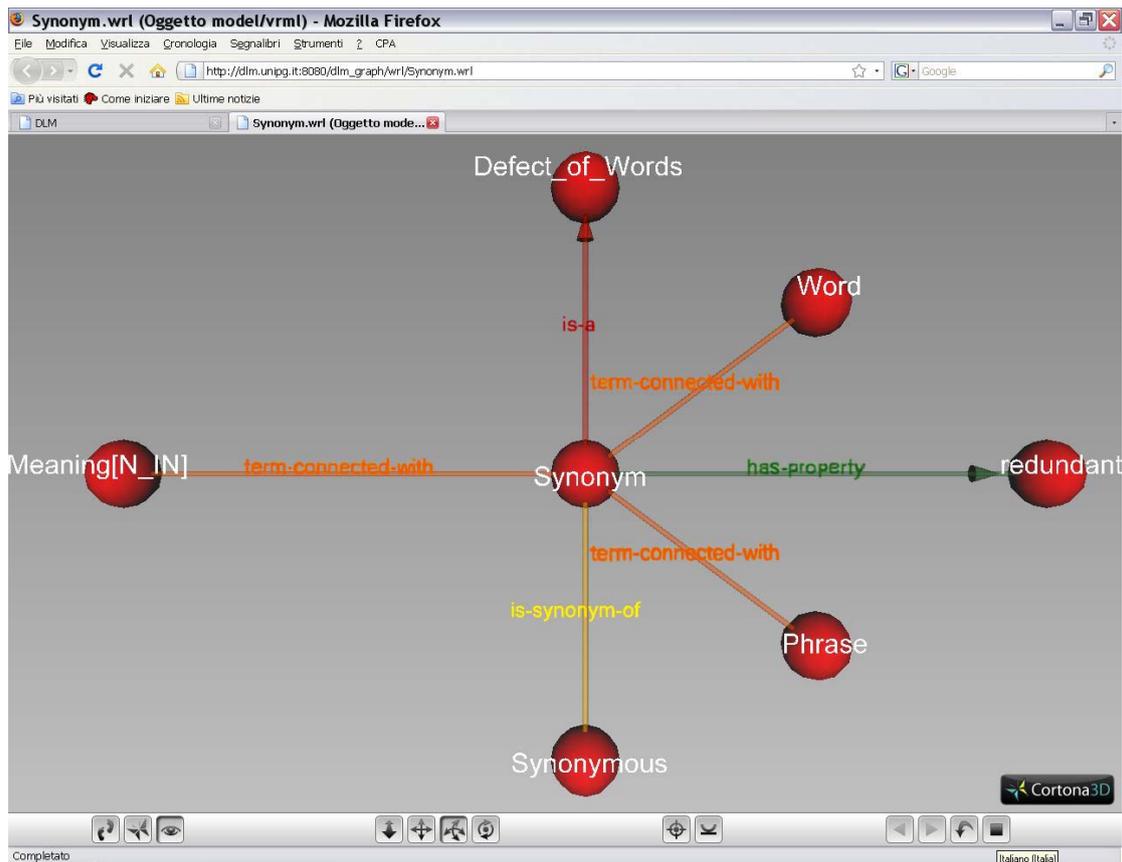

Figure 3: Graph of the of the term "synonym".

---

[20] "Of DISCOURSE, Or the several notions belonging to *Grammar* or *Logick* […]. THE most general name for those external expressions, whereby men do make known their thoughts to one another, is DISCOURSE The several things and notions belonging to discourse, may be distributed into such as do concern the *Parts of it*; or those primary ingredients of which it consists, *Less simple*; WORDS. II *That which is intended by any such sound or Character*, is called MEANING. Sense, signification, purport, acception, import, tenor, denote, moral" (Wilkins 1668:44, 46).

5 Conclusions

This survey has focused on the DLM as a case study for the consideration of the way in which the specialised knowledge that is organised in texts can be represented in LSP dictionaries. Concept defining properties are variously distributed in texts, and often remain implicit. In the DLM this problem is particularly evident because its definitions are directly drawn from author quotations. The initial model of the DLM provided terminological information in a bi-dimensional scheme but, in a more recent phase, it has been integrated with a relational chart which can point out the connections among the recorded terms. In this way the DLM is endowed with a multidimensional defining paradigm.

The examples drawn from the *Essay* that I have considered above suggest how the interrelation of the two sections planned for the DLM can help the user to have a more complete view of the concepts recorded in this on-line dictionary. In the blueprint of the reference work under examination a crucial role is attributed to the conceptual information as it is provided by the authors of outstanding linguistics texts. It is recorded in 'defining quotations', but these do not always give a clear-cut view of concepts – and we might also add that concepts do not always have a clear-cut outline within texts.

In the DLM entries the indication of synonyms and generic cross-references between terms is made available, but a more complete and detailed view of conceptual relations is guaranteed by their graphic representation. In the graph term relations are made explicit, also from a visual point of view. In the DLM, graphs simplify and, at the same time, also integrate the information that is provided both by the quotations and the sections of the text that are analysed by the compiler but cannot be included as explicit information in the records of the dictionary.

The complementary data contained in each single system are available to the user as a result of the interrelation between the two sections of the DLM: in these interrelating components the knowledge distributed in specialised texts is collected and organised while also maintaining its original format.

References

Antia, Bassey Edem (2000). *Terminology and Language Planning: an alternative framework of discourse and practice*. Amsterdam/Philadelphia: Benjamins.

Baumann, Klaus-Dieter (2001a). Der Fachtext als komplexes Wissenssystem. Ein interdiszi-plinäres Konzept. *LSP and Professional Communication* 1/2:8-33.

Baumann, Klaus-Dieter (2001b). LSP text as a knowledge system. In Mayer 2001:105-113.

Baumann, Klaus-Dieter (1992). *Integrative Fachtextlinguistik*. Tübingen: Narr.

Bergenholtz, Henning & Sven Tarp (eds.) (1995). *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam/Philadelphia: Benjamins.

Bloomfield, Leonard (1935). *Language*. London: George Allen & Unwin.

Brekke, Magnar (2001). LSP lexicography and terminography: a complementary view. In Mayer 2001:179-187.

Cabré, Maria Teresa (1999). *Terminology. Theory, methods and application*. Amsterdam/Philadelphia: Benjamins.

Cabré, Maria Teresa (1996). Terminology today, in Harold Somers (ed.) *Terminology, LSP and Translation. Studies in language engineering in honour of Juan Carlos Sager*, Amsterdam/Philadelphia: Benjamins, 1996:15-33.

Crystal, David (1997). The Linguistic Identity of English-language Dictionaries of Linguistics. *Lexicographica*, 13:17-33.

Crystal, David (1996). *A dictionary of linguistics and phonetics*. Oxford: Blackwell, 4th edition.

De Meo, Anna (2002). Lemmi e definizioni nel Dizionario generale plurilingue del Lessico Metalinguistico. In Lorenzi 2002:35-55.

De Meo, Anna & Franco Lorenzi (2006). Terminologia e lessicografia delle scienze del linguaggio: il Dizionario generale e plurilingue del Lessico Metalinguistico. In *Atti del Convegno "La Formazione in Terminologia" Portico di Romagna, 29-30 aprile 2005*. *Mediazioni* (DOI 10.1473 /media 36), Daniela Di Fidio (ed.), http://www.mediazionionline.it/english/special_issue/demeo-lorenzi_eng.html. (consulted January 2009).

Dik, Simon C. (1989). *The theory of functional grammar*. Part I: T*he structure of the clause*. Dordrecht [u.a.]: Foris.

Dolezal, Fredric T. (ed.) (1989). The dictionary as text. (Thematic issue) *International journal of Lexicography* 2/3:167-276.

Geentjens, Sancho *et al.*, (2006). *Sociocognitive terminology and Termontography*. http://cvc.ehb.be/pub/JETALA2006_SKRP.pdf (consulted January 2009).

Glossary of terms used in terminology. Bruno de Bessé, Blaise Nkwenti-Azeh & Juan C. Sager. (1997). *Terminology*, 4/1:117-156.

Gotti, Maurizio (2003). *Specialized Discourse: Linguistic Features and Changing Conventions*. Bern: Peter Lang.

Hüllen, Werner (2004). *A History of Roget's Thesaurus. Origins, Development, and Design*. Oxford: Oxford University Press.

Hüllen, Werner (1999). *English dictionaries 800-1700. The topical tradition*. Oxford: Clarendon Press.

Jansen, Hanne (2003). *Densità informativa: tre parametri linguistico-testuali. Uno studio contrastivo inter- ed intralinguistico*. Copenhagen: Museum Tusculanum Press.

Leonardi, Natascia 2008. John Wilkins' metalinguistic lexicon in the panorama of linguistic terminology. In Marijke Mooijaart & Marijke van der Wal (eds.). *Yesterday's words. Contemporary, current and future lexicography*. Newcastle: Cambridge Scholars Publishing, 2008:160-172.

Lorenzi, Franco (ed.) (2002). *DLM: Dizionario generale plurilingue del Lessico Metalin-guistico*. Roma: Il Calamo.

Maat, Jaap (2004). *Philosophical languages in the seventeenth century: Dalgarno, Wilkins, Leibniz*. Dordrecht etc.: Kluwer.

Mayer, Felix (ed.) (2001). *Language for special purposes: perspectives for the new millennium. Proceedings of the 12th European symposium on LSP* (2 vols.). Vol. 1. *Linguistics and cognitive aspects, knowledge representation and computational linguistics, terminology, lexicography and didactics*. Tübingen: Narr.

Ogden, Charles K. & Richards, Ivor A. (1923). *The meaning of meaning: a study of the influence of language upon thought and of the science of symbolism*. London: Routledge & Kegan Paul.

Rothkegel, Annely (2005). Knowledge and text types. In Helle V. Dam, Jan Engberg & Heidrun Gerzymisch-Arbogast (eds). *Knowledge Systems and Translation*. Berlin/New York: Mouton de Gruyter, 2005:83-102.

Saussure, Ferdinand de (1916). *Cours de linguistique générale* [Charles Bally & Albert Sechehaye (eds.)] Lausanne/Paris: Payot.

Schaeder, Burkhard & Henning Bergenholtz (eds.) (1994). *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: Narr.

Temmerman, Rita (2000). *towards new ways of terminology description. The sociocognitive approach*. Amsterdam/Philadelphia: Benjamins.

Wiegand, Herbert Ernst (ed.) (1990). *Dictionaries and their parts as texts*, (Thematic part) *Lexicographica International Annual for lexicography* 6:1-161.

Wilkins, John (1668). *An Essay Towards a Real Character and a Philosophical Language. By John Wilkins D.D. Dean of Ripon, And Fellow of the Royal Society. London, Printed for Sa: Gellibrand, and for John Martin Printer to the Royal Society*.

Wright, Sue Ellen (1997). Representations of concept systems. In Sue Ellen Wright, & Gerhard Budin (eds.) *Handbook of terminology Management* (2 vols.), Vol. 1. *Basic aspects of terminology management*. Amsterdam/Philadelphia: Benjamins, 1997:89-97.

***