Towards a filtering of the relevant semantic information from MRDs

Toni BADIA and Roser SAURÍ, Barcelona, Spain

Abstract

Machine-readable dictionaries (henceforth MRDs) are a valuable source of lexical data for the automatic construction of lexical knowledge bases (LKBs). However, the transfer of information from one source to the other one often demands a filtering process, since in many cases the original dictionaries contain mistakes, redundancies and granularity irregularities. We claim that this filtering process has to be linguistically oriented and, to this end, we start from an amended Generative Lexicon (GL) as a framework for the representation and treatment of lexical information. The use of such a frame allows us to distinguish two main kinds of polysemy.

1 Introduction

As is widely acknowledged, MRDs are a valuable source of lexical data for the automatic construction of LKBs. In many cases, however, the original dictionaries contain mistakes, redundancies and granularity irregularities. In order, then, to obtain reasonable NLP lexica a previous filtering step seems to be required. In this paper we claim that this filtering process has to be linguistically oriented. To this end we start from an amended Generative Lexicon (GL) as a framework for the representation and treatment of lexical information.

In the following section we relate our approach to the basic work in lexical acquisition from MRDs. Section 3 presents a general overview of GL and its developments, and justifies its usefulness as a theoretical model for the filtering process. In section 4 we analyse some lexical entries from our reference Catalan MRD. Finally the last section draws our main conclusions.

2 Settings

One of the main issues in Computational Lexicography is the construction of informative, coherent, and economic lexicons to be used in NLP applications. The need for these resources has emerged because of the birth of so-called corpus linguistics, which provides with large amounts of data to be used as input for NLP, and the subsequent abandonment of the "toy tools" conception.

Constructing a computational lexicon from scratch is an expensive and time-consuming task. Taking advantage of pre-existing lexical resources is a well-known approach within Computational Lexicography. Thus, automatic acquisition of lexical knowledge from MRDs has been an important issue within this field since the beginnings of the 80s (see for instance [Amsler 1980]). A great deal of the interest within this area focuses on the acquisition of semantic information, which in MRDs is mainly encoded in an implicit way. One of the goals of the work carried out in semantic information acquisition is the construction of Lexical Knowledge Bases (LKB);

i.e., lexical resources where the content, possibly derived from previous lexical data bases, is organised from a semantic, rather than from a superficial point of view [Rigau 1998:29]. A notable project within this line was Acquilex[1], which produced interesting work on automatic methodologies for the construction of LKBs from MRDs [Vossen 1992; Rigau 1998]. Other fruitful work in a similar direction was that carried out at Microsoft, which also led to the establishment of a taxonomy for lexical organisation [Dolan *et al.* 1993]. It is worthwhile to note, however, that there are other projects involving lexical organisation from a semantic point of view, although they start out from a diverging perspective: the CYC project [Lenat/Guha 1990], for instance, which has been manually developed following an introspective approach to lexical knowledge; or the widely known and exploited WordNet [Miller *et al.* 1990], which is a semantic net presumably based on psychological principles, rather than on the information stated in any dictionary.

Our concern here is the construction of a coherent and adequate LKB from a linguistic perspective. We are specifically interested in the organisation of the semantic information in the lexicon so that it can be used in real NLP applications. As stated in the literature, the systematicity (or unsystematicity) of the lexicographic sources used for the acquisition process conditions the adequacy of the semantic information and the level of soundness of the resulting taxonomy. Thus, the results obtained from a very consistent source such as LDOCE (Longman Dictionary of Contemporary English) [Procter *et al.* 1987], where the structure of the information was accurately designed from the very beginning, are considerably different from the ones obtained from other, less normalised lexicographic sources. Indeed, this second sort of MRD is the most common one, including the Catalan dictionary which is our starting point in this research [DIEC 1995]. Some recurring problems are circular definitions, the appropriate delimitation of different senses of the same word, and the question of sense granularity –that is, the presence of too fine-grained distinctions, not necessary in NLP applications. Furthermore, building lexical LKBs from MRDs results basically in a sense enumeration lexicon [Pustejovsky/Boguraev 1994]; that is, in spite of the hierarchical structure of the lexicon, this methodology results in a lexical organisation where it is not possible to represent the implicit or explicit relationships that exist between different senses of the same word, as well as between different words.

Although there is some research that has tried to avoid some of these problems (such as [Rigau 1998]), it has been carried out mainly using stochastic approaches. This means that from a theoretical point of view not enough attention has been paid to the notion of polisemy, to the differentiation of senses, and to the virtual relationships between words. To avoid producing a computational lexicon that inherits all of the MRD with its inconsistencies and mistakes, we need a means of filtering the semantic information that is exported from the MRD. This calls for a theoretical conception of the lexicon, a perspective that has to be developed within a linguistic-oriented framework.

The question then is how this filtering process can be performed. Note that there is no external criterion that can be used in this task, since a pre-defined list of acceptable senses to be exported would already imply that the solution is known. It is only by analysing the lexical entry in the MRD and discovering its internal structure that we may decide which senses should be exported on their own and which ones can be collapsed into a single one. Of course this can be done only if a theory of lexical meaning is available that makes it possible to distinguish theoretically between different senses of a word.

3 Building a theory of lexical senses

In move in this direction we propose to start from the model of semantic meaning that is being developed in the Generative Lexicon theory [Pustejovsky 1991; 1995], which builds on classical sources stemming from Aristotle. In GL lexical meaning is composed of two elements:

- a sound and coherent organisation of the semantic information in the lexical entries, with distinct layers of information, combined with a hierarchical conception of the lexicon, and
- a set of relations that different aspects of the meaning of words may have with one another.

The first element is the one that is most known and used in relation to GL. There are three distinct layers of semantic information, corresponding to the argument structure, the event structure and the qualia structure (with its quadripartite structure of four different qualia: the formal quale, the constitutive quale, the agentive quale and the telic quale), as explained in [Pustejovsky 1991; 1995]. To this basic organisation of the semantic structure of words, which we may call the standard GL representation, a few modifications have been proposed in the last few years. Roughly in [Copestake 1993, Copestake/Briscoe 1996], and more explicitly in [Badia/Saurí 1998; 1999] an integration is proposed of GL semantic representation and HPSG syntax, so that a reasonably straightforward interaction between syntax and semantics is available. In [Johnston 1996] the notion of a head quale within the qualia structure is introduced in order to distinguish the specific quale that is relevant in determining the referential properties of the word. In [Badia/Saurí 1998] the index of the HPSG semantic structure is maintained, again in order to fix referential properties. Finally in [Badia/Saurí 1999] the notion of prominence is applied to the qualia, so that a single quale can be highlighted in the respective pairs of the functional qualia (i.e.: the agentive and the telic) and the individuating qualia (i.e.: the formal and the constitutive). As may be noted all these modifications are oriented towards two basic aims: to develop the interaction of the semantic representation with syntax and to facilitate the selection of the preferred reading in each relevant context. Below there is a typical GL semantic description that contains all of these elements.

$$\begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{head} \\ \text{VALENCE} & \textit{lists-of-complements} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \textit{index} \\ \text{EVSTR} & \textit{event-structure} \\ \text{ARGSTR} & \textit{argument-structure} \\ \text{QUALIA} & \begin{bmatrix} \text{FORMAL} & \textit{formal-relation} \\ \text{CONSTITUTIVE} & \textit{constitutive-relation} \\ \text{AGENTIVE} & \textit{agentive-relation} \\ \text{TELIC} & \textit{telic-relation} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 1: Schema of an HPSG/GL-based lexical entry

This structure is easy to implement, because it is simply representational. That is to say there is no commitment as to how it affects the real processing of texts. As a matter of fact, a simplified version of the standard GL representation has been used in two EU-funded projects: Acquilex and SIMPLE[2]. Common to these approaches is that they do not implement the second aspect of GL, but only use its representational structure.

The second element of GL (the coding of the set of relations that different aspects of the meaning of words may have with one another) is also known but less used, probably because it is not so easy to implement: for example the LKB that was used to implement the Acquilex proposals [Copestake 1993] was not powerful enough to introduce the generative mechanisms implied in standard GL. However there currently are ways of implementing the generative capacity of the lexicon: basically this is dependent on a proper type system, with multiple inheritance and enough inference capacity. If these requirements are met, either with underspecification [Markantonatou/Sadler 1998] or default inheritance [Copestake/Briscoe 1992], or with both [Lascarides/Copestake 1999], most of the relations originally foreseen in GL (and some others) can be implemented. Implementations like the new LKB [Copestake 1998] allow much of what is needed.

Note that the generative capacity of the system is essential in order to reduce the number of lexical entries, and consequently to reduce ambiguity in analysis. If we obtained a representational system that allowed us to nicely express the different aspects of word meaning but with no way to relate those aspects of meaning, we would end up with a sense enumeration lexicon –more sophisticated, but still sense enumeration.

Our aim, then, is to use GL as a general theory of lexical semantics in order to find the kind of sense distinctions that are useful in NLP. To this end we foresee a processing of lexical entries in an MRD, in order to finally derive lexical entries for a computational dictionary. This filtering process would then be applied as part of the transfer of information from the MRD to the LKB. The task we envisage, however, is not that of going through a lexical entry of a dictionary for human use and to detect what part of the information it contains goes to what layer in the GL representation.[3] Rather, it is that of going through a lexical entry and detecting what is common in the different senses of that entry and to see whether the differences between them are explainable with the mechanisms we have at hand. In order to do this automatically, we would rely on the regularity of lexicographic definitions, for example, by using the possible patterns in definitions that concern each qualia role as indicators of the nature of the different pieces of information appearing in each definition.

4 Analysis of some lexical entries

In this section we examine some lexical entries considered to be semantically ambiguous in the reference dictionary. Our lexicographic resource is the most recent normative Catalan dictionary: the DIEC (1995). Although we work on Catalan data, our treatment can be easily transported to other languages with the appropriate reformulations which mainly concern the definitions patterns relative to the different qualia roles.

We first analyse the different senses of the chosen lexical entries. We then put forward a formal representation, following the lines introduced in the previous section, and consider what the

adequate mechanisms are to cope with polysemy from a generative approach.

• ***construcció* ('building')** Let us start with process-result nominalisations such as *construcció* ('building'). The DIEC presents the following main senses:

1. *Acció de construir.* ('Act of building.')
2. *Manera d'estar construïda una cosa. ex: Un edifici d'una construcció elegant.* ('The way something is built. ex: An elegantly built building.')[4]
3. *Cosa construïda.* ('A built thing.')
4. *Art o tècnica de construir.* ('Art or technique of building.')

These four senses are clearly related. Sense 4 is a fine distinction from sense 1: considering that *art* (sense 4 genus) is defined as 'system of rules and precepts to perform something well', sense 4 could be reformulated as 'act of building following a system of rules and precepts'. And this minor distinction between sense 1 and 4 appears to be unnecessary in NLP applications. Similarly, the result-denoting reading (sense 3) expresses a consequence of the process of building (sense 1) [Badia/Saurí 1998]. This can be observed in some contexts where the two readings (process and result) need to be considered, such as the example of sense 2 in the DIEC: here *construcció* expresses the event originating the edifice, which corresponds to sense 1, and simultaneously *construcció* also refers to the result, due to the presence of the state-denoting adjective *elegant*. Other examples where a reference to both the process and the result are made can be found with redescription predicates[5], such as *traduir* and *traducció*. In (1), for instance, *natural* is only recoverable as the result of the translating act.

(1) a. La traducció d' aquest pamflet m' ha costat molt, però al final crec
The translation of this pamphlet me has costed a-lot, but at-the end believe(1-s)
que m' ha quedat molt natural
that me has resulted very natural

b. Traduir aquest pamflet m' ha costat molt, pero al final crec que
To-translate this pamphlet me has costed a-lot, but at-the end beleive(1-s) that
m' ha quedat molt natural
me has resulted very natural

Therefore, the different senses of *construcció* in the DIEC should be collapsed into a unique lexical entry expressing a relationship between the process and the result readings. The entry as is stated below is adequate to represent *construcció* (and similar predicates) in the contexts where both the process and the result are implied. Note however that the head value in the event-structure acts as a mechanism allowing for the selection of one of the two main senses (process and result) when appropriate.

$$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \textit{index} \\ \text{EVSTR} & \begin{bmatrix} \text{EVENT1} & \text{e1: } \textit{process} \\ \text{EVENT2} & \text{e2: } \textit{state} \\ \text{RESTR} & \text{e1} \langle \text{e2} \\ \text{HEAD} & \text{e1/e2} \end{bmatrix} \\ \text{ARGSTR} & \begin{bmatrix} \text{D-ARG1} & \text{x: } \textit{animate-ind} \\ \text{D-ARG2} & \text{y: } \textit{artifact} \\ \text{D-ARG3} & \text{z: } \textit{material} \end{bmatrix} \\ \text{QUALIA} & \begin{bmatrix} \text{FORM} & \text{exist}\left(\text{e2,y}\right) \\ \text{AGEN} & \text{constr-act}\left(\text{e1,x,y}\right) \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 2: Lexical entry for *construcció* ('construction')

• ***pota* ('paw')** A different case of polysemy is the one illustrated by *pota* ('paw'). the DIEC enumerates two basic senses:

1. *Qualsevol dels membres d'un animal que li serveixen per a aguantar el cos i per a caminar, córrer o saltar.* ('Any part of an animal serving to hold the body and to walk, to run and to jump.')

2. *En els mobles i altres objectes que en llur part inferior acaben en tres o més barres amb què recolzen al soler, cadascuna d'aquestes barres.* ('In furniture and in other objects having three or more vertical pieces serving to support the object on the floor, each one of these sticks.')

Here the common elements of the two senses are: the description of the entity as an extremity, the relationship that this mantains with a second individual (which is of different sort in each case), and finally a general aspect of the utility of the entity referred to. These three different aspects are related to the formal, constitutive and telic qualia roles, respectively.

In the feature structure below we use distributed disjunction [Krieger/Nerbonne 1993] to explain the link between the alternates of the two disjunctions in the qualia structure: in the formal and in the telic quale. This is not problematic in this paper since we only explicitely describe the representational part of the semantics of words. However, in a real system other means may be used that are probably more adequate: e.g. default multiple inheritance [Lascarides/Copestake 1999].

$$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \end{bmatrix} & \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \text{x} & \\ \text{EVSTR} & \textit{evstr} & \\ \text{ARGSTR} & \begin{bmatrix} \text{ARG1} & \text{x} \\ \text{ARG2} & \text{y: \$1 } \textit{animal} \text{ V \$2 } \textit{piece\_of\_furniture} \end{bmatrix} & \\ \text{QUALIA} & \begin{bmatrix} \text{FORM} & \text{extremity}\left(\text{x}\right) \\ \text{CONSTIT} & \text{part\_of}\left(\text{x,y}\right) \\ \text{TELIC} & \text{hold}\left(\text{x,y}\right)\text{\& \$1 move\_with}\left(\text{y,x}\right) \end{bmatrix} & \end{bmatrix} \end{bmatrix}$$

Figure 2: Lexical entry for *pota* ('paw')

• ***forca* ('fork')** There are some other cases that behave similarly to *pota*, in the sense that they present distinct senses related by common information in one or several qualia roles. An example of an entry with distinct senses due to a change in the telos is *forca* ('fork'):

1. *Pal amb dues o tres puntes o branques en un extrem que serveix per a regirar, apilotar, etc., palla, fems, etc., agafar i carregar garbes, etc.* ('Long handle with two or three prongs at the end, used to move and make piles, etc., of straw or dung, etc., and to take or load bundles of wheat, etc.')

2. *Pal acabat en un extrem en dues puntes emprat pels forners per ficar llenya dins el forn.* ('Long handle ending with two prongs used by bakers to put wood inside an oven.')

3. *Barra de fusta que per un extrem és bifurcada, emprada pels vidriers per agafar i transportar peces que tinguin coll.* ('Long handle with two prongs, used by glassworkers to take hold of and to transport glass pieces with long necks.')

$$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \text{x} \\ \text{EVSTR} & \textit{evstr} \\ \text{ARGSTR} & \begin{bmatrix} \text{ARG1} & \text{x: } \textit{instrument} \\ \text{ARG2} & \text{y: } \textit{artifact} \\ \text{ARG3} & \text{z: } \textit{artifact} \end{bmatrix} \\ \text{QUALIA} & \begin{bmatrix} \text{FORM} & \text{x} \\ \text{CONSTIT} & \text{handle}(\text{y,x}) \,\&\, \text{extremity}(\text{z,x}) \,\&\, \text{exist\_2}(\text{z}) \\ \text{TELIC} & \text{manipulate}(\text{w,v,x}) \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 3: Lexical entry for *forca* ('fork')

These three senses present an almost identical description of the form of the object and are distinguished by the telicity of each particular instrument.[6]

• ***tanca* ('fence')** Other entries present distinct senses due to a change in the individuating qualia (that is, the formal and constitutive roles). This is the case of *tanca* ('fence').

1. *Reixa, estacada, bardissa, etc., amb què se circumda un espai de terreny, amb què es tanca una obertura.* ('Grate, grill, railings used to enclose a piece of land or some other opened area.')
2. *Qualsevol tipus de separació de terrenys, d'habitatges o de propietats feta de fusta.* ('Any type of border between pieces of land, houses or wooded properties.')
3. *Paret de pedra seca que circueix una parcel.la.* ('Stone wall surrounding a property.')
4. *Balda, barra, etc. amb què es tanca una porta, un finestró; dispositiu amb què es tanca una bossa, un collaret, etc.* ('Bolt or bar used to lock a door, a window; or any device used to close a bag, or to lock a necklace, etc.')
5. *Bloc d'acer que serveix per a obturar la recambra de les armes de foc, en especial la de les peces d'artilleria.* ('Steel bolt used to fix the chamber of a fire arm, especially in artillery weapons.')

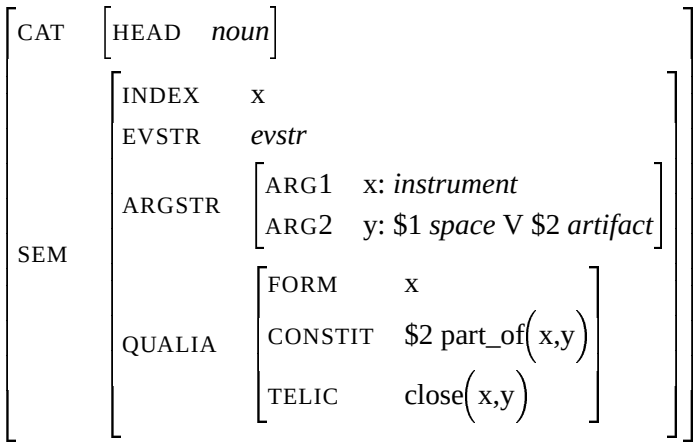

Figure 4: Lexical entry for *tanca* ('fence')

The first three senses are very similar. They have exactly the same telos and differ only with regard to the form or constitution. The fourth one, by contrast, is only loosely related to these ones and just shares a general telos with them.[7] Finally, the fifth reading is similar to the fourth one. All five senses can be roughly collapsed to two: the first three on the one hand and the last two on the other. The observable difference between them is that in the latter *tanca* constitutes a part of an artifact –actually, the artifact on which it carries out its function.

5 Conclusions

The transfer of information from MRDs to LKBs often demands a filtering process. After analysing several lexical entries and proposing a representation for them within a unification-based LKB, we have shown how the semantic structure proposed in GL (specifically, the qualia roles informing the qualia structure) helps in filtering the different senses stated in lexicographic definitions.

In addition, we have seen that a flexible and generative lexicon requires the use of dynamic mechanisms that can account for the different semantic realisation of words in context. Two main kinds of polysemy have been distinguished here. On the one hand, there are cases such as *construcció*, which require a unique lexical entry where the several facets of their meaning have to be represented. As shown, the formal mechanism applied here works on the syntagmatic axis, in the sense that takes advantage of the selectional restriccions imposed by the elements in context. On the other hand, there are cases such as *pota, tanca*, and *forca*, which present different senses that are related by the sharing of information in one or more qualia levels. In contrast with the previous kind of polysemy, for these cases we have established a representation of the semantic core of the word, to which the possible specialisations of the word are added in a compact way. This representational format, which can be implemented in a default multiple inheritance system, guarantees the proper inferences through the LKB hierarchy when a particular sense is required.

Two sorts of conclusions can be drawn from this exercise: a first one concerning the general requirements of NLP applications, and a second one of a lexicographical nature.

First, when importing semantic information from dictionaries for human use with the purpose of building LKBs, there is a need to examine the distinctions that they establish between senses, so that unnecessary distinctions are not carried over; particularly those senses linked to distinctions that are general enough to be considered in the grammar. In many cases a general sense description is sufficient, provided that the system enables inference between types, and general operations for generating new types are allowed.[8]

Second, dictionaries for human use (at least the DIEC) are often not well constructed: users/readers are sometimes misled, and the basic meaning of a word is not explicitly given. The way in which word meanings are treated wrongly suggests that they are fixed once and for all, and that new extensions are not possible. This is particularly important in societies/cultures where normative and standardisation dictionaries are seen as legislators of the words that are licensed or correct in the language. The work put forward in this paper shows that a more flexible approach should be taken when defining lexical meanings, thus allowing for the relation of different instantiations of what is in reality one single meaning. GL provides the theoretical elements and mechanical devices for such an approach.

Notes

[1]Esprit-BRA 3030 and Esprit BRA 7315

[2]http://guagua.echo.lu/langeng/projects/simple/index.html

[3]As in the projects mentioned above.

[4]The introduction of examples will not be systematic through this paper, but only when necessary for the sake of the explanation.

[5]They differ from the standard creation predicates in that, when used with a process sense, they cannot take an argument that describes the result of that process.

[6]Another sense has to be added in Catalan: that of the killing instrument ('gibbet' in English). Originally it had the same form as the other instruments denoted by *forca* but, of course, with a different telicity –just another example of the same phenomenon. These days the dictionary only describes more modern forms of this killing instrument, which do not resemble the original one very much in form.

[7]Note that the differences (in form and in telicity) included in this fourth sense are very similar to those that justify the first three senses.

[8]For real applications what might be needed is a specification of the sub-languages in which that word has a specific use.

References

[Amsler, R. (1980)] Amsler, R. (1980) *The Structure of the Merrian Webster Pocket Dictionary.* Ph.D.diss., University of Texas, Austin.

[Badia, T. & R. Saurí (1998)] Badia, T. & R. Saurí (1998) 'The Representation of Syntactically Unexpressed Complements to Nouns'. *COLING-ACL'98. Workshop on the Computational Treatment of Nominals,* p: 1-10. Montréal.

[Badia, T. & R. Saurí (1999)] Badia, T. & R. Saurí (1999) 'Semantic disambiguation of adjectives in local context: a generative approach'. In P. Bouillon & E. Viegas *Description des Adjectifs pour les Traitements Informatiques.* Workshop. TALN'99. Corsica.

[Castellón, I. (1992)] Castellón, I. (1992) *Lexicografía computacional: adquisición automática de conocimiento léxico.* PhD thesis, Universitat de Barcelona.

[Copestake, A. (1993)] Copestake, A. (1993) *The Compleat LKB*. Acquilex-II deliverable, 3.1. ms. CCL.

[Copestake, A. (1998)] Copestake, A. (1998) *The (New) LKB system*. http://hypatia.stanford.edu/ aac/lkb.html (updated: Dec 10,1998).

[Copestake, A. & T. Briscoe (1992)] Copestake, A. & T. Briscoe (1992) 'Lexical operations in a unification-based framework'. Pustejovsky, J. & S. Bergler (eds.)*Lexical Semantics and Knowledge Representation. Proceedings of the ACL SIGLEX Workshop on Lexical Semantics and Knowledge Representation.* Berkeley. Berlin, Springler-Verlag: 109-119.

[Copestake, A. & T. Briscoe (1996)] Copestake, A. & T. Briscoe (1996) 'Semi-productive polysemy and sense extension'. In Pustejovsky, J. & B. Boguraev (eds.) *Lexical Semantics. The problem of polysemy*. Oxford, Oxford University Press.

[Copestake, A., A. Sanfilippo, T. Briscoe & V. de Paiva (1993)] Copestake, A., A. Sanfilippo, T. Briscoe & V. de Paiva (1993) 'The Acquilex LKB: An introduction'. In Briscoe, de Paiva & Copestake (eds.) *Inheritance, Defaults and the Lexicon.* Cambridge University Press: 148-163.

[DIEC (1995)] DIEC (1995): *Diccionari de la Llengua Catalana*. Institut d'Estudis Catalans.

[Dolan, W., L. Vanderwende & S. Richardson (1993)] Dolan, W., L. Vanderwende & S. Richardson (1993) 'Automatically deriving structured knowledge bases from on-line dictionaries. *Proceedings of the first Conference of the Pacific Association for Computational Linguistics (pacling'93).* April 21-24, Simon Frase University, Vancouver, Canada.

[Johnston, M. (1996)] Johnston, M. (1996) 'Semantic underspecification in lexical types: capturing polysemy without lexical rules'. *Acquilex Workshop on Lexical Rules, 1995.* Cambridge.

[Krieger, H. & Nerbonne, J. (1993)] Krieger, H. & Nerbonne, J. (1993) 'Feature-based inheritance networks for computational lexicons'. In Briscoe, de Paiva & Copestake (eds.) *Inheritance, Defaults and the Lexicon.* Cambridge University Press.

[Lascarides, A. & A. Copestake (1999)] Lascarides, A. & A. Copestake (1999) 'Default Representation in Constraint-based Frameworks'. *Computationla Linguistics,* vol. 25, 1: 55-105.

[Lenat, D.B. & R.V. Guha (1990)] Lenat, D.B. & R.V. Guha (1990) *Building Large Knowledge-based Systems: Representation and Inference in the Cyc Project.* Addison Wesley.

[Markantonatou, S. & L. Sadler (1998)] Markantonatou, S. & L. Sadler (1998) 'Lexical generalisations'. Van Eynde & Schmidt (eds.) *Linguistic specifications for typed feature structure formalisms.* Luxembourg, EC.

[Miller, G. (1990)] Miller, G. (1990) *Five papers on WordNet.* Special Issue of International Journal of Lexicography, 3 (4) 1990.

[Procter, P. et al.] Procter, P. et al. (eds.) (1987) *Longman Dictionary of Contemporary English* Longman, Harlow and London.

[Pustejovsky, J. & B. Boguraev (1994)] Pustejovsky, J. & B. Boguraev (1994) 'A Richer Characterization of Dictionary Entries: The Role of Knowledge Representation'. In: Atkins & Zampolli (eds.) *Computational Approaches to the Lexicon*, Oxford University Press.

[Rigau, G. (1998)] Rigau, G. (1998) *Automatic Acquisition of Lexical knowledge from MRDs*. PhD thesis, Barcelona, Universitat Politècnica de Catalunya.

[Vossen, P. (1992)] Vossen, P. (1992) 'The Automatic construction of a knowledge from dictionaries: a combination of techniques'. In H. Tommola *et al.* (eds.) *5th Euralex International Congress on Lexicography*. Tampere, Finland. pp. 311-326.