Extraction of Knowledge about Terms from Indications of Metalinguistic Activity in Texts

Carlos Rodríguez Penagos

A description of metalinguistic fragments in sociological research papers, has led us to suggest Explicit Metalinguistic Operations (EMOs) a notion which refers to textual contexts that are rich in knowledge about the rules, conditions and elements of linguistic code. We claim that EMOs have unique recurring formal traits that allow for its exploitation in terminological and lexicographic work. We propose a method for automatically detecting contexts of term creation and modification.

1 INTRODUCTION

The intention of this paper is two-fold. First, we discuss some of the linguistic and philosophical issues arising from metalinguistic phenomena. We focus particulary on statements about language, which constitute vital nodes of knowledge in specialized texts. We argue that these coding and decoding operations have deep and lasting repercussions in the constitution of common linguistic frameworks for the construction of a knowledge validated by consensus within an expert community. Secondly, we will also report on some empirical work done for the domain of academic research. We suggest that using knowledge about how metalinguistic statements work can allow the retrieval of semantic and pragmatic information on terms from unprocessed text.

In section 2 of this paper we discuss the importance of metalinguistic statements for linguistic knowledge extraction and representation. Section 3 is a brief review of the main characteristics of metalinguistic activity in texts, and of the vital role it plays at intersubjective knowledge-building. On section 4 we introduce the concept of *Explicit Metalinguistic Operations* and present empirical work done on a corpus of academic articles from peer-reviewed research journals. In section 5 we discuss the implementation of these principles into applications that can be of help for lexicography and terminology. Finally, section 6 will offer some conclusions and point to future work.

2 SETTINGS

It is an obvious fact that the retrieval of any kind of useful knowledge from text has to be proceed from a framework of linguistic analysis and interpretation.

What is more difficult to say is at what level of complexity this analysis should be attempted, and where should we best invest our processing resources in order to detect those sections of text that are important or relevant for our purposes. Until recently, NLP research developed approaches that privileged those aspects of language that more easily adapted to logic and formal constraints, and so were better suited for algorithmic treatment. Syntactic patterns or grammatical classes could, for example, be useful to provide candidates for term detection. On the semantic side, observed regularities of meaning were more difficult to formalize though they could claim to be the actual repositories of knowledge about the world. In fields such as Automatic Translation, semantics was sometimes restricted to queries sent to machine-readable dictionaries (MRD) or other implicit means of semantic representation, such as conceptual hierarchies or ontologies which are ruled by meronymic and hyperonymic relationships.

But we already have a powerful device for storing and transmitting human knowledge. From as early as Aristotelian times definitions were identified as discourse devices able to transmit and create knowledge, both linguistic knowledge (when we define a word) and encyclopedic knowledge about the word when we define a concept that supposedly reflects how reality is organized. By the late 70s Amsler & White [1979] were already extracting *IS_A* links from the relationship between a defined term and the syntactic head of dictionary definitions in order to create taxonomical structures. Recent efforts along similar lines by Dolan, William *et alia* [1993] at Microsoft aimed at building a structured lexical knowledge base. The Acquilex project aimed at the extraction of lexical information from machine-readable dictionaries, giving rise to a huge amount of valuable research. Kruijff & Schaake [1995] have studied ways to establish relevance using informational structure (Topic-Focus) in order to help extract definitions from text. More recent efforts ([Pascual & Péry-Woodley 1997], [Pearson 1998], or [Cartier 1998]) have attempted automatic extraction of knowledge from specialized texts, such as software manuals or textbooks, by analyzing relatively fixed and stable definitional patterns that are delimited (sometimes excessively) by formal and structural constraints.

What many of these approaches lacked was a way to extract pertinent semantic data from sentences that did not always adhere to the usual formalism of didactic or lexicographic definitions. Structure and form in sentences supplying information about language is much more heterogeneous than the classic *word = genus + differentia* schema. An extraction method that can only account for clear-cut definitions would miss many other fragments of text where relevant information is being produced that refers to the way language is expressing the knowledge we have about our world. That knowledge is being created within a community, and needs to be transmitted in order to be validated. That can only be done through linguistic interaction, through rational consensus. Statements about language, whether in a formal system or in Natural Language, are vital nodes in the

interaction between content and form, between knowledge and expression, between personal intuition and intersubjective scientific understanding. We will now discuss some basic aspects of metalinguistic predication in Natural Language before proceeding to describe their role in highly specialized texts.

3 METALANGUAGES AND THE FOUNDATIONS OF EXPERT KNOWLEDGE

3.1 Metalinguistic issues in NLP

Reflexivity (the property of referring to itself) has been ascribed to language as one of its most important features ([Hjelmslev 1943], [Tarski 1944], [Carnap 1934], [Jakobson 1963], and [Rey-Debove 1978], to name but a few), and one that sets it apart from other semiotic systems. In order to do this mirroring, language has to be split (at least methodologically) into two distinct systems that share the same rules and elements: a metalanguage is a language that is used to talk about another one (an object language), which in turn can refer to and describe objects in the world. Metalanguages are foundational in nature [Lara 1989], as they have the power not only of describing a communicative code, but of directly enacting and creating it. A metalanguage supplies the framework on which a linguistic code can actually mean anything at all, defining on the one hand the formal elements that belong to it and, on the other, the combinatory rules allowed for in the construction of meaning and sense in a well-formed sentence. A metalinguistic predication thus establishes the conventionality of meaning first put forward by Saussure that enacts a linguistic code system, and establishes specific signs as elements capable of conveying significance or sense.

A discontinuity occurs in semiotic levels when linguistic items are "mentioned" instead of being "used" normally in an utterance. This notion can explain what happens when the word "Socrates" does not refer to a person, but to itself as linguistic sign, e.g. in the classic example:

(1) "Socrates" is an eight-lettered word.

or account for a change in the grammatical properties of verbs such as "moving" or "suffering" when used in a metalinguistic context:

(2) moving means changing position in space through time

(3) « Suffering » **is a** mild **word** to describe what I felt then…

Coseriu [1986] states that any element of linguistic code can become a name unto itself ([Rey-Debove 1978] and [Carnap 1934] call them"autonyms") and automatically acquire nominal traits, in what Rey-Debove terms a "metalinguistic rewriting rule". When a metalinguistic statement is embedded in regular statements of an object language, this is usually signaled prominently using various expressive means: with lexical items that act as descriptors (e.g., *term* or *word*) or metalinguistic verbs (*called, termed, dubbed, etc.*), with recurrent syntactic structures, and other pragmatic or paralinguistic resources (hedging, typographical conventions, layout on the physical page or punctuation). A metalinguistic sentence has to be *flagged* somehow in order for the decoder to be able to infer the special meaning and conditions attached to it.[1] The decoder has to provide the adequate context for the succesful interpretation of the utterance. Markers like quotation marks or descriptors force the interpretation within a framework of metalinguistic activity of the items under their influence, automatically making them acquire the typology "*X is_a_sign*".

3.2 Theories and the build-up of language

Theoretical and computational linguists have always been acutely aware of the central role that an effective formalism plays in an adequate theory of how language works.. A realization of the semiotics of model-building has become fundamental in the way our discipline views itself as an empirical domain.

> "(...) the predictive power of the theory arises from the conventional correspondence between the model and the empirical domain." [Pollard & Sag 1994]

It is precisely the creation of “conventional correspondence” that is being carried out under metalinguistic control, as noted as early as 1957 by Roman Jakobson. In theoretical research, knowledge is constantly being created and monitored through metalinguistic statements that allow for negotiation of the conceptual frameworks that are materialized in complex terminologies. Non-specialized language exchanges are not abundant in this kind of discourse operations because (unless in the context of language-acquisition) we usually rely on a lexical competence that, although it can be constantly modified and enhanced, reaches the plateau of a common lexicon relatively early in our adult life. Nonetheless, when we want to structure and acquire new knowledge we have to go through a resource-costly process of integration within coherent conceptual structures of a considerable quantity of new and very complex lexical items or terms. Robust knowledge representation formalisms have to efficiently capture the dynamic, aggregative and metastable nature of expert knowledge.

---

[1] Nirenburg & Levin [1991] specify that in order to avoid ambiguity a metalanguage for the description of natural language should not consist of lexical units of the same language.

3.3 Metalinguistic activity in *expert-to-expert* communication

Unlike the relatively stable lexicon of everyday language, specialized terms are continually being created, put forward or modified purposefully by an expert community in order to accomplish their communicational and representational aims. The *technification* of meaning can be understood as an abstract representation of empirical phenomena [Wignell 1998], which in the realm of lexical knowledge results in the need for putting forward terms which have very specific meanings and usage conditions constantly negotiated and dependent on the consensus of an expert community or an academic group.[2] Terms are the means to *objectify* reality in order to allow for its cognitive manipulation. Metalinguistic statements in highly specialized texts thus enact the sense or usage specificity (with regards to a posited general lexical competence) which constitutes the very technical nature of terms, while at the same time facilitating the interpretation and cognitive processing of scientific discourse.

This terminological control is actually being done in well-bounded textual fragments that usually serve to state something about the value, meaning and/or usage conditions of the lexical items that are focused in a metalinguistic statement. Definitions of the kind employed by dictionary makers, where an hyperonymic genus is linked up with specifying *diferentiae* in order to conceptually establish a word within the framework of language (or to constraint it to a technical domain) are just one of the ways in which metalinguistic operations can materialize in texts. Nevertheless, on many occasions the kind and scope of the information retrieved does not respond to an inflexible paradigm in which hyponymy, meronymy and conceptual completeness rule. Many times what really is being provided is partial information; for instance, just an additional semantic trait for an item already described, or the modification of a pragmatic restriction, or maybe the writer is merely suggesting a synonym or even attempting an evaluation of the lexical item when in a restricted context.

4 EXPLICIT METALINGUISTIC OPERATIONS IN A SPECIALIZED CORPUS

In general, definitions and other metalinguistic operations can be viewed ([Jakobson 1963], [Riegel 1987], [Kleiber 1990], [Bierwisch & Kiefer 1969]) as deep-structure equations that relate a term with its semantic content, or as answers to a lexical question such as: "¿what is the <meaning | usage conditions | referent | value> of the linguistic sign X?", that carry out the transformation of language by explicit modification of the lexicon. As stated above (section 3)

[2] "In those cases that conceptualizations deviate from the knowledge we have, expert knowledge on the object obviously has to use different words, constructions or specialized senses to be able to refer to this knowledge." [Meijs and Vossen 1991]

such discourse operations have to be somehow marked or flagged in order to be succesful.

We have suggested elsewhere [Rodríguez 1999a and 1999b] that **Explicit Metalinguistic Operations** (or **EMOs**, as they will be referred to from now on) are useful notions when dealing with the wide variety of surface structures that such statements about language can take. We analyzed a corpus of 19 sociology articles published during the last 5 years in various British, American and Canadian academic journals with strict peer-review policies. In 138,183 words, 151 EMOs constituted a 3% of total, which although statistically sparse does not accurately reflect their epistemic importance. It is in this thin slice of data that language is accomplishing a fundamental intersubjective task of building new knowledge and testing the old one. That is why it is vital to precisely tune our retrieval tools to find that (and only that) solid and valuable information.

For the sake of clarity we now present two examples of EMOs from our corpus, as well as examples from catalan an spanish corpora: [3]

> (4) Integral power results in a fundamental type of social classification which, adapting Bernstein's terminology, I shall call "frame" (Bernstein 1971).
>
> (5) The bit sequences representing quanta of knowledge will be called "KENES", a neologism intentionally similar to "genes".
>
> (6) En este entramado se puede llegar a distinguir sucesivas expresiones de lo heredado, aunque el término fenotipo se suele reservar para la expresión final.
>
> (7) La fàcies, amb l'embotornament causat per edema palpebral, presenta un aspecte clínic característic que s'ha anomenat "fàcies xarampion-oide".

In order to identify such operations in ongoing text, we have specified some minimal requirements and some basic elements associated with them that should always appear in the surface structure of the sentences actually performing such operations. The features in listing I describe the general requirements that a sentence or a phrase should have in order to be considered an EMO.

Listing I. EMO requirements:

i) The presence of **a linguistic sign that is the subject** (either logical or grammatical) of a predication that needs not be a complete sentence.

ii) The utterance should count as a contribution of relevant **information about the status, coding or interpretation** of a lexical unit.

---

[3] Corpus Crea (RAE): http://www.rae.es/NIVEL2/recursos.htm and Corpus textual especialitzat plurilingüe IULA/UPF: http://www.iula.upf.es/corpus/ corpus.htm, respectively.

iii) The whole operation should be **prominent or conspicuous on account of its metalinguistic nature**.

An EMO, then, should always have at least three basic constitutive elements corresponding to these requirements, which we present in listing II. Following each one, we show how each of those elements can be projected on an example sentence (5):

Listing II. Constitutive elements of EMO and projection on example sentence

A) a (complex or simple) term or linguistic sign functioning as **a self-referential term** or **autonym**, which stands as the logical or grammatical subject of the metalinguistic operation;

KENES

B) **a semantic or pragmatic content** to be linked up with the term or lexical unit;

The bit sequences representing quanta of knowledge
KENES **intentionally similar to "genes"**.

C) a metalinguistic, or "definitional" verb or VP, a typographical, punctuation or dispositional mark or a combination of other semiotic resources that **connect (and conceptually articulate) element A and element B**, **while flagging** the extra-ordinary (non-referential) nature of the segment (**Markers/operators**).[4]

The bit sequences representing quanta of knowledge **will be called "** KENES **", a neologism** intentionally similar to "genes".

In order to better envisage how each element is distributed in any sentence, we have adopted a parenthetical markup that accomplishes just that:

(5) / <The bit sequences representing quanta of knowledge> [will be called "]KENES[", a neologism] <{intentionally} similar to 'genes'>. /

(8) The new world order could be / [']SYNCRETIC[',] [a term] <from the study of religion> {(see Colpe, 1987).} /

where: angular brackets **< >** signal the semantic or pragmatic content of the EMO, that is, the actual information supplied for the lexical item, here highlighted with UPPERCASE letters. Square brackets **[ ]** show the markers or operators that articulate and make salient such operations, and the last set of brackets { } help identify information we could term "peripheric", or encyclopedic [Bierwisch & Kiefer, *ibid*], that do not belong to the core of semantic or lexical knowledge of the domain. The slashes / ... / spotlight the boundaries of the operation, spanning the whole sentence in (5), and just a segment of it in (8).

[4] We group these functions together because sometimes they are materialized with the same linguistic and ortographic elements, though they should be separated from a logical and theoretical point of view.

We suggest to reduce the diversity of EMOs to two very basic discourse actions: to **inform** on a (linguistic) state of affairs (that is, establish and present a standing communicative code), or **direct** either the coding or the decoding of a message. Both can be seen as being equivalent on a higher level of abstraction, as an informative statement can imply that a convention stands, and could be interpreted as a hidden directive to follow when communicating within a group.

In our previous work we have also suggested [Rodríguez 1999b] a tentative typology of EMOs which we will only quickly review here, as well as an open list of other potentially extractable information that could be pertinent for terminology or for computational lexicography. This typology allows us to assign to each kind of EMO unique features that are not derivable from linguistic form alone, but have to do with their nature as discourse objects. We have observed, though, that most of the time EMOs are mixed (that is, they are, for example, informing on an state of language and at the same time supplying a referent through denotation).

Listing III. Tentative typology for EMOs:
I. Informative:

a) **Denotation:** contributing information that identifies the referent (*extensional* dimension), or modifies the meaning of a lexical unit (*intensional* dimension), so as to be distinct from others in a terminological system.

b) **Connotation:** contributing information pertinent (but not essential) to the identification of the terminological unit insofar an element of the specialized linguistic code. Here we could include judgements about value, indicators of source (as in citations) or pragmatic restrictions.

c) **Equivalence:** supplying a relation of functional equivalence or equisignificance between two lexical units (for example, synonymy), without perhaps offering concrete semantic characteristics.

II. Directive or instructional:

They can be either *interpretative* (that is, pertinent for decodification within a certain context) or *directive* (concerning pragmatic restrictions or conditions for linguistic coding).

Besides the actual materialization of a semantic or pragmatic equation, EMOs can supply some additional information pertinent for the exchange and structuring (through its linguistic expression) of specialised knowledge. These expressive devices help the speaker position himself with regard the actual knowledge being evaluated, as well as in the consensus processes taking place within the expert group. In listing IV we present other aspects that can be potentially retrieved from some metalinguistic statements.

Listing IV. Other pertinent information retrievable from EMOs

- **Extent or scope of the proposal:** Local (applies to present text), regional (for a specific theory or problem) or global (valid for a whole domain).
- **Participants** in the communicative exchange.
- **Locutionary force.**
- **Attribution of semantic responsibility:** who is putting forward the new term, where and when.
- **Attitude of speaker** (towards others or towards his own utterances).

In the following listing (V) we present an open list of elements that appear in sentences where EMOs are present. The nature of each one contributes in their own special way to the overall process. Some of the information that we can obtain for terminology or lexicography is encoded in the choice of elements that a writer actually uses (or avoids using). The use of one descriptor or another, *name* vs. *term,* for example, is actually providing additional information about the proper classification of the linguistic item being considered. As we have emphasized, orthographic and dispositional elements introduce important aspects of metalinguistic activity that can be easily detected. Our findings suggest that metalinguistic activity is a discourse process in which many levels of linguistic resources interact. At the same time, the actual set of recurring items that operate those processes is limited and can in principle be recorded exhaustively.

Listing V. Common recurring elements in EMOs

A) Lexical**:**

- Descriptors: *term, word, phrase, terminology, vocabulary, name,* etc.
- Metalinguistic verbs: *calls (is called), means, termed, name, refer, use (is used), speak, designates, known as, stands for, defined as, coins, corresponds to, dubbed, designates, labels, indicate, said (to be),* etc.
- Other lexical indicators: *subtitle, oxymoron, where* (in formulas), etc.

B) Syntactic: Apposition, copulative clauses, etc.

C) Pragmatic: Informational structure, hedging, etc.

D) Paralinguistic resources: Layout (footnotes, highlighted text, tables, etc), typography and punctuation.

There are two important observations to be made here. First of all, none of these elements by itself is enough for a positive identification of an EMO. It is the whole articulation of items from different dimensions of linguistic and textual structure that bring about a complex, foundational discourse action such as the ones we have been analyzing. These elements and patterns are not always exclusive to metalinguistic statements, but most of the time they can signal that such processes are taking place in a text. Nevertheless, the presence of any element as

indicator of autonymy has to be reinforced by other elements (either formal or semantic) in that operation, as some of those items are polysemous or can perform different functions in different contexts; for example so-called "scare quotes" can merely indicate tentativity, or some otherwise reliable lexical indexes of metalinguistic activity could be doing something else completely, as in the following example from the *Brown* Corpus:

> (9) In any case it is by no **means** clear that formally structured organs of participation are what **is called** for at all .

A second trait of EMOs relevant for its computational exploitation concern the fact that they can be materialized inter- or intra-sententially: with anaphoric links to complete sentences or expressions, as in example (10), or embedded in a *guest/host* relationship (adapted from [Leech 1980]) within a complete sentence which is not metalinguistic, as in example (11):

> (10) ...<This> [shall be referred to as] the FLUX PATTERN.

> (11) Demographic checks on $_{\text{host}}$ / $_{\text{guest}}$ what has become known as THE MISSING MILLION $_{\text{guest}}$ / $_{\text{host}}$ are described by…

5 IMPLEMENTATION ISSUES

In what follows, we will discuss the use of EMOs in NLP applications. As we have stated above, the inventory of elements and patterns that signal and enact metalinguistic activity is quite limited (at least in the specialised contexts we have examined). Consequently, we do not need complex syntactic analysis or large lexicons in order to detect those important discourse operations. The task at hand is to fine-tune a finite-state device or a regular expression application that can discriminate such elements without being distracted by similar, non-metalinguistic, sentences like (9). A more exhaustive inventory of distinctive elements than the one we have obtained from our sociology corpus should be attempted. The indicators we supply here would have to be complex in order to allow for the inclusion of phrasal elements, clauses and for their "remote control" coordination (when acting inter-sententially) with elements from other non-lexical dimensions, such as typography or punctuation. In contrast to previous proposals, we suggest to incorporate into the processing some indicators and indexes that are not lexical in nature. This means that added complexity is introduced by the consideration of different levels of linguistic description, but it also means that a more richly-textured information is made available to lexicographers and terminologists.

We conceive of Explicit Metalinguistic Operations as knowledge-rich contexts that can be retreived and processed using devices[5] that first become aware of a metalinguistic segment in text by noticing in it some of the elements from listing V, and then go on to determine the lexical item that is occurring autonymically following rule-encoded heuristics. The next step would be to assign the role of semantic-pragmatic content to some other text strings within that sentence or sentences.

Lexical, grammatical and paralinguistic indicators of metalinguistic activity such as the ones we have presented above could trigger the recognition and retrieval of EMOs from unprocessed text.[6] Rules that encode observed regularities could then identify three kinds of elements, marking surface elements as **markers/operators (**OP/MRK**), terms** (TERM) or **content strings** (CONT). Although we have not yet implemented such a system, we anticipate that the processing could be semi-automatic, with the system presenting candidate terms and contexts to a human reviewer who will have the final say on how to structure and store the lexical and terminological information, if it is found to be relevant for the task at hand.

Which kind of relationship between a sign and its meaning or referent, is obtainable from these processes could be inferred from the interpretation of the semantics and pragmatics of each of the finite set of OP/MRK elements (e.g. *"I shall call"*, in example 4) and of the kind of specific elements that are identified as CONT. In this way we could reliably have access to information about its unique place in a coding system, or about its usage conditions or restrictions. This shallow metalinguistic parsing could use a finite state grammar that is not theory-dependent and is easily adapted to scanning a text for pragmatic, lexical, syntactical and paralinguistic features.

Information thus retrieved could be transferred to simple terminological Knowledge Base fields as entries for TERM, for OP/MRK, and finally for a content field (CONT) that stores the expressions which constitute the actual semantic or pragmatic information. This heterogeneous information bank could be enhanced and modified subsequently by processing larger and more exhaustive corpora. Those latter additions should not simply overwrite previous ones, so as to preserve a record of the process that has led to consolidation.

In order to illustrate such processing, we show next how a sample sentence ends up dissasembled in a terminological record that stores its pertinent information.

---

[5] Similar to the Contextual Exploration Method put forward by Desclés [1997], and implemented for Spanish by Couto, Crispino, et alia. [1999]. There are other low-level systems for detecting terms that use, for example, Constraint Grammar [Voutilainen 1993], but none (as far as I known) provide an explicit treatment of metalinguistic information.

[6] As each language uses its own specific expressive resources for making sense and for constructing meaning, inventories of idiosyncratic indexes and markers for each one (maybe even domain-dependent ones) could be compiled.

Example sentence (4) parenthetically marked

> Integral power results in / <a fundamental type of social classification which>, {adapting Bernstein's [terminology], I} [shall call] [']FRAME['] {(Bernstein 1971)}/.

A possible record resulting from the decomposition and processing of our example sentence could be constructed along the following lines, where each element of the EMO is assigned a field in the data base:

Field	**Record**
Term	*Frame*
Operators/markers	" " | *terminology* | *call*
Semantic/Pragmatic content	*a fundamental type of social classification (integral power results in)*
Other available information	*Semantic responsibility:* I (writer) *Source:* Bernstein_1971 *Scope:* local

Some conceptual and lexical inferences could then be made available for enriching, complementing and constructing a more exhaustive and systematic dictionary entry, or for its use by AI applications capable of exploiting them:

1) "Frame" *(as lexical item) is_a* term
2) Frame *(as concept) is_a* fundamental type of social classification
3) a fundamental type of social classification (integral power results in) *is_called* Frame
4) Writer (I) *Proposes* "Frame" from Bernstein_1971

6 CONCLUSIONS AND FUTURE WORK

We have shown the importance we assign to metalinguistic activity in texts, both for NLP and for processing highly-specialized texts. We have also suggested that some recurring formal and contextual features in certain metalinguistic segments that we have termed Explicit Metalinguistic Operations can be used for automatically detecting term creation and modification without requiring large processing resources. Information thus obtained can be very useful for lexicographic or terminological applications.

But our proposal is just one way of using knowledge about how metalinguistic activity proceeds cognitively and linguistically, and we are sure other applications could be envisaged. We only claim here that the indicators we have found

point to conceptually important nodes of text, and that they are not just collocational data with no epistemic importance. Even though we in no way claim to have provided one, we do suggest the need for applications that can "mimic" our impressive human competence as efficient readers of technical subjects, as incredibly good lexical-data processors that constantly update and construct our own special purpose vocabularies. In the end, this paper is just an invitation to look beyond the notion of conventional definitions and exploit the more general and powerful dimension of metalinguistic activity in the search for better and smarter processing of language and knowledge.

REFERENCES

Amsler, R. & White, J. (1979) *Development of a computational methodology for deriving natural semantic structures via analysis of machine-readable dictionaries.* Technical Report TR MCS77-01315, L. R. C., U. of Texas.

Bierwisch M. & Kiefer F. (1969) *Remarks on definitions in general language*. Studies in Syntax and Semantics. D. Riedel Humanities Press, NY

Carnap, R. (1934) *The Logical Syntax of Language.* Routledge and Kegan, Londres 1964.

Cartier, E. (1998) *Analyse Automatique des textes: l'example des informations définitoires.* RIFRA'98. Sfax, Tunisia.

Coseriu, E. (1967) Teoría del Lenguaje y Lingüística General. *("Sistema, norma y habla, 1952")* Gredos, Madrid.

Cuouto, J., Crispino et al. (1999) *Estructuración de Índices Gramaticales y Léxicos para la Extracción y Recuperación de Información.* Procesamiento del Lenguaje Natural. SEPLN. No. 25

Desclés, J-P. Et al. (1997) *Textual Processing and Contextual Exploration Method.* In CONTEXT'97, Río de Janeiro, Brasil

Dolan, William., et al. (1993). *Automatically Deriving Structured Knowledge Base from On-line Dictionaries.* Proc. of the Pacific Association for Computational Linguistics, 1993, Vancouver, British Columbia.

Hjelmslev, L. (1943) *Prolegomenos a una teoría del lenguaje.* Gredos, 1974

Jakobson, R. (1957) *El Metalenguaje como problema lingüístico.* El Marco del Lenguaje. FCE, México 1988

Jakobson, R. (1963) *Ensayos de Lingüística General.* Ariel, Barcelona. 1984.

Kruijff, G. & Schaake, J. (1995) *Discerning Relevant information in discourses using TFA.* Fifth CLIN Meeting (Twente, November 23 1994).

Lara, L. F. (1989) *Une critique du concept de metalangage.* FOLIA LINGÜÍSTICA, Acta societatis Linguistica Europaeae. Tomus XXIII /3-4, Mouton/De Gruyter

Leech, G. (1980) *Explorations in semantics and pragmatics.* Amsterdam, John Benjamins

Meijs, W. and Vossen, P. (1991): *In so many Words: Knowledge as a Lexical Phenomenon.* In Lexical semantics and Knowledge Representation

Nirenburg, S. & Levin, L. (1991) *Syntax-driven and Ontology-driven Lexical Semantics*. Lexical semantics and Knowledge Representation 1st. Siglex Workshop; Pustejovsky-Bergler, eds. Springer-Verlag.

Pascual, E. & Péry-Woodley, M-P (1997) *Modèles de texte pour la définition.* Journées Scientifiques et Techniques du Réseau Francophone de l'Ingénierie de la Langue, Avignon

Pearson, J. (1998) *Terms in Context*, John Benjamins (Studies in Corpus Linguistics) Vol 1., Amsterdam

Pollard, C.& Sag, I. (1995) *Head-Driven Phrase Structure Grammar,* University of Chicago Press.

Riegel, M. (1987) *Definition directe et indirecte dans le langage ordinaire: les énoncés définitoires copulatifs.* Langue Francaise 73.

Rodríguez, C. (1999a) *EMOs in specialised discourse: the construction of lexical meaning in theoretic science.* Terminology and Knowledge Engineering 1999, Innsbruck, Austria.

Rodríguez, C. (1999b) *Operaciones Metalingüísticas Explícitas en textos de especialidad. Treball de Recerca.* Institut Universitari de Lingüística Aplicada. Pompeu Fabra University.

Tarski, A. (1944) *La concepción semántica de la verdad y los fundamentos de la semántica.* Nueva Visión, Buenos Aires. 1972

Voutilainen (1993) *NPTool, a detector of English noun phrases.* Proceedings of Workshop on Very Large Corpora. Ohio State University.

Wignell, P. (1998) *Technicality and abstraction in social science*. Reading Science: Critical and functional perspective on discourses of science. Martin, J.R. & Robert Veel, Routledge

Carlos Rodríguez Penagos is a doctoral student at the Institut Universitari de Lingüística Aplicada (IULA) of the Universitat Pompeu Fabra at Barcelona. La Rambla 30-32, 08002 Barcelona. carlos.rodriguez@iula.upf.es