

THE CODIFICATION OF ONTOLOGICAL AND CONTEXTUAL INFORMATION FOR PRECOMPUTATIONAL TREATMENT IN GRAMMARS AND ONTOLOGIES

LA CODIFICACIÓN DE INFORMACIÓN ONTOLÓGICA Y CONTEXTUAL PARA EL TRATAMIENTO PRECOMPUTACIONAL DE GRAMÁTICAS Y ONTOLOGÍAS

*Margarita Goded Rambaud**

ABSTRACT

One problem at the linguistic preprocessing stage has to do with the concepts included in existing linguistic models. Part of the problem of codifying ontological and contextual information focuses on the lack of differentiation between communication and cognition that some linguistic models present. Besides, there are some described linguistic concepts that are lightly marked and which lack enough empirical textual, lexical or grammatical evidence that support them.

Because a unified linguistic model able to account for ontological and contextual information is not yet available, a simpler mechanism capturing linguistic, ontological and contextual information can be simpler at a preprocessing stage. Instead of using whole linguistic models, it is explained here how an algorithm describing the components that make up linguistic codification can be used to facilitate precomputational codification. This algorithm is based on the structural similarity of the grammar of a language, the ontology supporting it and the proper descriptive algorithm. Finally, the use of this algorithm illustrates how to extract this information from a corpus.

Key words: Precomputational codification, ontologies, dictionaries, descriptive algorithm, lightly codified meanings.

RESUMEN

Uno de los problemas del preprocesamiento del lenguaje tiene que ver con los conceptos que se incluyen en los modelos de descripción lingüística existentes. Una parte del problema de codificar la información ontológica y contextual incide en la carencia de diferenciación entre las funciones de comunicación y de cognición que presentan algunos modelos lingüísticos. Además, hay conceptos lingüísticos descritos como conceptos débilmente marcados que carecen de suficiente evidencia empírica textual, léxica o gramatical que los soporte.

Al no existir todavía un modelo unificado de descripción lingüística que sea capaz de dar cuenta de las restricciones de codificación ontológica y contextual, es posible proponer un mecanismo de captura de la información lingüística, ontológica y contextual más simple y utilizable en un estadio precomputacional del procesamiento del lenguaje. Así, en vez de usar modelos lingüísticos completos para una codificación precomputacional, se explica aquí cómo un algoritmo de descripción de componentes que intervienen la codificación lingüística

* Universidad Nacional de Educación a Distancia. Profesora Titular de Semántica. España.
Correo electrónico: margarita.goded@flog.uned.es
Recepción: 15/1/2016. Aceptación: 16/3/2016

puede facilitar esta parte del proceso. Este algoritmo se basa en la equivalencia estructural de las gramáticas de las distintas lenguas, las ontologías que las soportan y el propio algoritmo de descripción. Finalmente se describe cómo usar este algoritmo para extraer esta información de un corpus.

Palabras clave: Codificación precomputacional, ontologías, diccionarios, algoritmo descriptivo, significados ligeramente codificados.

1. Introduction

In this paper, the text extraction of ontological and contextual information for precomputational treatment is explored together with some problems related to the codification of lightly codified meanings (LCM) for NLP applications. Firstly, the concept of LCM is presented. Then alternative theories for their treatment are discussed in connection with the role that context play in them, and finally, a procedure to handle such lightly linguistically codified meanings for their precomputational treatment in the form of a descriptive algorithm is explained and examples of its application are shown.

Communication and information are related but they are different concepts. In this paper, communication is defined as the process of transmitting information among humans and or machines. While the first focuses on the information, which is transmitted by either humans or machines, the second term refers to the proper semantic content, separate from the nature of its transmission. Also, the most immediate approach in the analysis of communication was initially related to language, but this is not, by all means, the only one.

Since human perception limits the way information is processed, and human cognition shapes the way it is handled, both approaches should be considered.

A basic preliminary question often posed among linguists and computational linguists alike is how the information interchanged among speakers is much richer than the information linguistically codified?

The problem underlying this work is why certain meanings, which are successfully

communicated among speakers and that have been extensively described in theoretical linguistics literature, are not fully machine understandable.

The basic idea is that relying on context retrieval, the amount of information communicated or retrieved among speakers is higher than the one that is linguistically codified. This specific linguistic codification is unnecessary because that information can be retrieved from other sources either ontological, encyclopedic or both. As a result, some part of information remains undercodified, or even it is left uncoded.

Linguistic codification has to do with the characteristics and limitations of the human cognition and communication systems. This insight was originally proposed by Nuyts (2004: 278) and expanded in Nuyts (2007: 554-55). The lack of differentiating these two functionalities is far reaching, and it challenges an important number of both functional and cognitive developments in present theoretical linguistics.

Non linguistically codified information is drawn from different sources, and it is then treated in some ways using different types of cognitive and logical operations such as deduction, inference, induction, etc. However, purely linguistic information is restricted to the specific limitations and characteristics of the human communication system that language is, and to the specific grammar of the particular language in each case.

If not linguistically, how all this information is codified to be communicated among speakers? How nonlinguistic information relates to proper linguistic one? To what extent, an analysis of these differences would help NLP developments?

The following implicatures are simple linguistic interchanges among humans of a variety of cultures, where kitchens and restaurants exist and which are perfectly understandable as basic underlying contextualizations.

- (a) Can I still have dinner?
- (b) I have already cleared the kitchen / packed up the kitchen.

Or the restaurant variation:

- (c) Can we still have dinner?
- (d) Sorry, the kitchen is closed.

These dialogues are clear cases of implicatures interpreted as NO by humans. However, a machine must be fed with other than strictly linguistic information from both, an ontology and further encyclopedic knowledge, to be able to draw the appropriate answer. An ontology of any kind or format should be able to link [dinner] and [kitchen] as part of human feeding contexts and thus as related lexical entries. Using conventional ontological and semantic automatic extraction analysis, it is possible to infer that, [a dinner is a time-scheduled type of human meal] [meals are sometimes prepared in household or restaurant places such as kitchens] [kitchens are areas that need to be cleared after been used and usually not used until the next feeding time]

So once the food preparation process is finished, the schedule allotted for dinners is over in both cases. In fact, a link to programs such the English Visual Thesaurus would probably solve this problem, with a lesser ontological and encyclopedic burden.

Alternatively, a probabilistic analysis of choices under the game theory as in Parikh (2010) is surely also a line to explore.

2. Lightly codified meanings (LCM)

Lightly codified meanings (LCM), as put forward by Goded (2012) about the concept of intersubjectivity as in Nuyts (2001a, 2001b, 2004, 2012), are certain meanings commonly transmitted among speakers but difficult or impossible to isolate in specific, exclusive and independent phonetic, lexical and morphosyntactic

codificational terms. Some of the LCM studied here are closely related to the sensory description of wines and have been extensively described either in the form of metaphors or from another linguistic perspective. The description of the wine tasting experience takes its prototypical form in the wine tasting notes written by oenologists or other wine experts.

These meanings are transmitted, not because they are linguistically codified in any proper distinctive conventional linguistic way, but because other than conventional linguistic information is being retrieved and other than linguistic rules have been in operation. The use of repositories of information in the form of ontologies and cognitive operational procedures and constructs are the obvious candidates. As Vossen (2003: 465) claims, because language is inherently ambiguous, "it relies on the shared knowledge we have to fill in in missing or distorted information". NLP technologies try to decode the minimal, ambiguous and implicit messages encoded in language, using only a fraction of all the background information that humans have.

Comparing classical ontologies with lattices, Vossen (2003: 465-469) explains "how language tends to lexicalize just those concepts that are efficient to support communication." Therefore, there are certain concepts that are not fully lexicalized, grammaticalized or pragmatically marked and that rely on shared background knowledge for its interpretation and transmission. Metaphorization and intersubjectivity, as cases of LCM, are analyzed in Robert Parker's 2007 wine tasting notes database and presented in Goded Rambaud (2012).

3. The no model approach

In Grisham (1986: 6), it was suggested that computational linguistics seems to be a good ground to test how linguistic theories work. It could be added that, given the need to strain the codificational aspects of language to make them machine understandable, theories are put against the ropes so that their weaknesses and strengths show up more easily.

In Goded Rambaud (2007), it was noted that, in contrast with other disciplines with a well-established body of agreed and shared basics of knowledge, linguistics, probably because as a separate discipline is still quite young, cannot claim to have such an agreed and shared body of linguistic knowledge. Thus, a unified theory of language is still to be formed in the field. Due to the present lack of it, a claim to use a sort of no-model approach was made.

In Butler (2012) it is suggested that a combination of ideas from different approaches could be much more powerful than any of the models he thoroughly compares taken by itself.

Saeed (2016: 354) claims that the two main trends in theoretical linguistics include the formal Chomskyan group of theories and a cluster of functional theories where cognitivism can be more naturally integrated. Although the generative trend, considered as a constellation of approaches, all of them within the Chomskian paradigm, has reached some internal consensus, it traditionally ignores basic considerations such as the fact that communication is the main functionality of a system such as language.

It should be noted that certain functional theories tend to include cognitive considerations in an attempt of comprehensiveness, without reflecting that the inclusion of cognitive operations of linguistic categorization in a linguistic model is not equal to describing a language operability. For instance, a cognitive perspective helps to explain the lexicalization of hyperonymy in certain cases, but it does not help with the identification of language specific working rules of grammars of particular languages. Let alone, if these theories claim for language typological adequacy.

Because cognition and communication are separate but related functionalities, with different operational units to start with, no linguistic model should be overloaded with functionalities belonging to other systems as first claimed in Nuyts (2004). While cognition allows the organization of perceived information in humans, linguistic communication codifies such information in a way that it can be transmitted by speakers. Both obviously interrelate. Language is

precisely where these two functionalities merge, interact or even overlap, but the way in which this merging, exploration or interaction takes place is still a question under study and debate. Besides, any analysis starting from the distinction of the different functionalities of communication and cognition also raises the issue of the extent to which a well-intentioned indiscriminate addition of these different functionalities would provide a more illuminating picture. One thing is claiming for a common core of ideas and quite another mixing up the functionalities of communication and cognition.

A linguistic model can only account for what can be linguistically documented. In this sense, Hengeveld and Mackenzie (2008) also restrict the effects of context to what have systematic influence in grammatical choices. Connolly (2007, 2011), contests this approach as if it were opposed to the use of collocations to obtain empirical evidence. In my view, both grammatical and collocational evidence are not necessarily incompatible.

The fact that gathering and analyzing a vast amount of data is now technologically possible opens the possibility of statistic approaches that can be interpreted from a variety of linguistic angles. Moreover, denying the possibility of launching creative theories, even if still without empirical evidence, will be tantamount to trying to limit human imagination. However, few claims without a sounded empirical support would go very far these days, where scientific advances are based on both theoretical soundness and empirical evidence.

In Goded Rambaud (2012), it was argued that one thing is claiming for a common core of identified and shared basic linguistic concepts and quite another mixing up the different functionalities of communication and cognition as claimed in Nuyts (2004). Thus, though advances along these lines are increasing, a unified theory of language is still to be formed in the field.

A combined approach should not be mistaken with attempts to include cognitive, functional or formal “parts” in the description of the linguistic system itself with the good

intention of reaching some basic “political” compromise between the functional, generative and cognitive paradigms. Sometimes it could be impossible to isolate phonetic, lexical, morphosyntactic and semantic information separately in terms of specific, exclusive and independent codificational terms. Maybe this more a question of understanding how language works in real life rather than trying to make it fit into any already existing well designed theory.

Some agreement seems to be emerging, though. It has to do with the ontological recognition of the structure of the predicate, presented under different formats, notations and emphasis. The concept of predicate frames, as the basic scaffolding concept, capturing the basic differences between entities and relations, is the key concept in Dik (1989, 1997) functional grammar and it has had important developments all affecting the nature of linguistic description.

As above mentioned, Hengeveld and Mackenzie (2008) stress the fact that relevant grammatical distinctions should be empirically identifiable in the first place, and Connolly (2011) fully integrates context in his model. Mairal Usón and P. Faber (2002), Perrián Pascual and Mairal Usón (2010), include a functional grammar knowledge base component in the construction of a computational environment, and Parikh (2010) proposes a logic-mathematical description of an equilibrium based model of language functioning. Furthermore, generativism has recognized this difference in its logic form, underlying the structure of the clause from the very beginning of the Chomskian formalizations. Cognitive Grammar as in Langacker (1991:283) recognizes how objects and interactions relate and how they are conceptually dependent. In Langacker (1987: 215), he explains that one cannot conceptualize interconnections without also conceptualizing the entities that are interconnected.

This basic agreement differentiating both basic concepts has had profound effects in the linguistic backgrounding of the present developments of ontologies. In computational terms, both entities and relations are frequently labeled as “objects” in the sense that, once they

are computationally translated, they are just “objects” subject to computational treatment.

So, although this paper doesn’t claim any particular faith or theoretical affiliation, the theoretical background this work draws from basically concerns a broad functional perspective and a distinction between the two main functionalities of language: communication, and cognition, as in Nuyts (2004). It also agrees with a basic idea that there is an equilibrium between the different types of information communicated by humans, as in Parikh (2007, 2010).

This specific non affiliation claim and the no-model approach are based on general agreements about basic ontological distinctions that seem to underly present theories of linguistics. Precisely because most theories seem to agree with these distinctions, a more theory independent approach is taken here. And because of this, the codification and the lack of it in certain lightly marked meanings is explored here, using only a precomputational tool in the form of an algorithm. In this no-model approach, the different types of linguistic and non linguistic information can be treated under basically any format and linguistic model, as long as there is some equilibrium among the components.

The object of this study has to do with finding ways in which certain concepts are formalized to be communicated, linguistically or other and the recognition of a lack of markedness as a default situation that needs to be codified as such. This requires that the roles of the lexicon and grammar in whatever the linguistic model used are clearly delimited so that the codification of relevant lexical and grammatical distinctions can be properly codified and the handling of overlappings identified from the beginning.

4. The codification of ontological and contextual information

The relevance of this type of information was first identified in Hymes, (1972), where what he defined as the setting included animate and inanimate entities present and time and space location. He also described the scene including participants with their social and psychological

properties and degrees of formality, adding a general extra mental context defined as multidimensional and hierarchical. This information is also accounted for in Connolly (2007), where the discourse/situation, physical / socio cultural, narrow / broad and mental / extra-mental types of contexts are described. It should also be noted that in Connolly (2011), a model of context called Extended Model of Context is developed, and he proposes the inclusion of probabilistic information.

Schalley and Zaefferer (2007) claim that progress in linguistic theorizing is increasingly informed by cross linguistic research and that comparison of languages relies on those concepts that are essentially the same across human minds, cultures, and languages and, that, therefore, can be activated through the use of any human language. They add that these instances of mental universals join other less common concepts to constitute a complex structure in our minds, a network of cross-connected conceptualizations of the phenomena that make up our world. They label such system of conceptualizations as ontology.

Also, they submit the idea that the most reliable basis for any cross-linguistic research lies in the common core of the different individual human ontologies.

An important development in such codification can be found in Niremburg and Raskin (2004). About the brevity aspect for codification in ontologies, Niremburg and Raskin (2004:166) explain how for computers brevity of the kind that they refer to, “has little real physical sense in these days of inexpensive storage, fast indexing and searching algorithms”. They also note that what is difficult for computer systems is “precisely making reliable and relevant inferences” and because of this they also claim in Niremburg and Raskin (*ibidem*) that, “spelling out as many inferences as possible from a text and recording them explicitly in a well-indexed manner for future retrieval is essential for supporting a variety of computational applications”. This is precisely what Parikh (2010) does when he identifies a combination of factors such as reference, use, indeterminacy, and equilibrium in linguistic codification.

4.1. The codification of ontological and contextual information in Parikh

Designing an ontology requires establishing an open list of entities [concepts] and a closed list of relations operating in such domain, semantic field or area of knowledge. Thus defining entities and relations is a previous requirement in the writing of any ontology, whatever its character.

Contextual information, on the other hand, can be defined by exclusion as any information that is transmitted by speakers of a language but that is not specifically codified by the grammatical and lexical means of this particular language.

In Parikh (2007: 921) he emphasizes the fact that “communication can be probabilistic (sic) rather than certain “ and in Parikh (2010:303-328), he also holds that the extraction of ontological and contextual information is based, among other things, on the probabilistic analysis of a processable utterance. Parikh’s treatment of context is preformalized regarding what is defined as “situationess”. His approach to meaning combines four distinct ideas into a single, unified framework including the concepts of reference, use, indeterminacy, and equilibrium.

Based on Grice’s (1957, 1969), rather functionalist stance, that the essence of use lies in communication and information flow between agents, Parikh (2010:13) expanded this exploration into an infinite lattice of concepts of communication and information flow, as well as related infinite lattices of concepts associated with speaker’s meaning and addressee interpretation, using game-theory definitions to explain these concepts. However, the mechanics of how such concepts are associated with the speaker’s meanings remains unclear, unless this information is extracted from either an ontology or an online encyclopedia or both.

Parikh’s objects of analysis are utterances, not sentences, which he connects with his model-theoretic structures, that is to the development of his “situation theory” where he provides a method for computing the interpretation of complex expressions in a non-Fregean way.

It could be interpreted that, because a Fregean way to deal with language is basically logic and Parikh is aware of the problems of trying to capture contextual information in a purely logical format, he developed the concept of “situatedness” to deal with this. This move leads him to take utterances as carrying much more contextual information than sentences and to use the game theory in the computation of infons. Because of this, his model is relevant for NLP developments. And also because his objects of analysis are utterances and not sentences, and because utterances involve choice and strategic interaction, game theory plays a fundamental role in his computation.

In Parikh (2007: 921), it is claimed that “communication can be probabilistic rather than certain”. It is also assumed that communication refers only to linguistic communication, while information also includes all other kinds of encyclopedic knowledge that speaker and addressee share, including sensory information. However, what seems to be surprising is Parikh’s (2010) tenet is that common knowledge enters only if one is analyzing Gricean communication, but if one is analyzing the wider notion of information flow, there is no need for common knowledge. Parikh (2007, 2010) also developed a Bayesian net model of how probabilities are generated in the context of anaphora resolution, but apart from in exemplary cases, likelihood assignment is not a question addressed there. It is also unclear how nonlinguistic encyclopedic information can be fed into the system so that likelihood can be assigned.

These different aspects of human language such as reference, use, indeterminacy, and equilibrium were used in the initial design of the DA in Goded Rambaud (2007), but the application of Parikh’s set of interactions in the DA is still an issue under development. Together with it, the question to be addressed is how encyclopedic knowledge interrelates with utterances in a way that probabilistic assignments can be made, and the correct inferences could be extracted.

In Parikh’s work, it is also claimed that equilibrium is just a computational technique to solve games, where game theory (and situation

theory) are merely the best current available tools to an approach to language and meaning defined as equilibrium semantics situation theory. What in Parikh (2010) is defined as infons, seems similar to a predicate structure (as made up of entities and relations): the logic-mathematical structures linking entities (concrete or abstract) and their relations.

From an epistemic perspective, in Parikh (2010: 39), agents observe situations first, identify some of the infons those situations contain next, and finally extract as uniformities some of the infons’ constituent relations and individuals. However, when these entities are described more formally, he proceeds in the inverse order: starting with individuals, properties, and relations, going on to infons and then defining situations.

Parikh (ibidem) holds that a constraint is a link among two situations that captures the essence of meaning. Because of this, constraints can be mathematically represented as (higher-order) infons. Such infons are defined (Parikh 2010: 38) as “individuals having properties and standing in relations”. That is, such relation involves at least two arguments for the linked situation.

This simply implies a very basic mathematical function such

$$f(x...n)$$

where f is an ontological function that relates one or more than one entities or a property of one or more entities. As explained and illustrated in Goded Rambaud (2015:47) and in Goded, Ibáñez and Hoste (2015: 210), a basic ontology codifies entities and relations and there are two types of basic relations. The predicate and its arguments relation and the entity and its properties relation. The lexicalization of relations is usually materialized in the form of verbs, adverbs, and adjectives in the languages that feature such grammatical categories and the lexicalization of entities takes the form of nouns and names in a nearly all human languages.

Parikh’s claim that much of the system of language and meaning can be expressed as a set of interacting constraints, supports the idea that the bottom line of most linguistic models ultimately

lies on the need to stipulate a series of entities of any kind and the relations that link them.

What interests us is how, when as in Parikh (ibidem), a “situation”, is defined (in plural) as “items of information.. which is true for at least one existing situation...and then called fact...”, such components can be isolated and handled. That is the detailed nature of the components of infons, their codificational nature and the procedure by which infons are inserted into a system in the first place. In other words how infons can be identified and described.

A well-known possibility emerges when using any linguistic parser that immediately identifies the syntactic subject. As a result, it identifies agency and the rest of the semantic components participating in this particular state of affairs. In addition, such simple parser, using the previously established phonetical, lexical, morphosyntactic, semantic and pragmatic identification markers of a given language, will be able to identify the logical and semantic structure of certain terms in a language, as in Goded Rambaud (2015: 165). This is interpreted here in terms that the infon involves the type of grammatical codification that this particular language recognizes as a rule of grammar of that language. For example, the codification of a double plural in English and its negation (both/nither) shows how the infon must include a type of link between two and only two entities or the lack of such double link. All this takes us back to the evaluation of linguistic models that can best capture the nature, possibilities, and limitations of infons.

Therefore, infons can also be interpreted as the base to the logical component which underlies most linguistic theories. From those theories within the Chomskian paradigm and its logical form to those in the functionalist paradigm such as (Dik (1997), Hengeveld and Mackencie (2008), Van Valin, et al.(1997), where various developments of the predicate structure form a cluster of converging theories.

There is also a traditional logical approach to Types in connection with the idea of intensionality, as in Kreidler (1998: 132): “the extension of a lexeme is the set of entities which

it denotes (...) the intension of any lexeme is the set of properties shared by all members of the extension.”. Thus, in formal semantic terms, the extension of a property in some situation is the collection of objects holding that property in that situation. However, in Parikh, this takes us back to a loop in his approach to intensionality because what Parikh concludes is different from the logical approach in the sense that his two types and extensions are, in fact, situated objects, being defined only regarding situations. The universe of entities interpreted from an information perspective is called the ontological or informational space (O).

Parick’s situation theory recognizes what is defined as arity of an infon (the number of argument roles an infon can take -unary, binary, tertiary infon, n-ary infon, etc.) where the argument roles have to be filled in by compatible objects. Parikh (2010: 46) holds that there is no general way to characterize this relation of compatibility that he takes as primitive.

This idea of mathematical valence applied to functions and predicates in predicate structures can also be identified in most linguistic models as a logical sub layer of meaning. So, although there are many other approaches, linguistic models subsuming different layers of ontological and logical codification are possibly the best candidates, precisely because they feature a sort of ‘by default’ logical and ontological blueprint easily translatable to machine languages.

4.2. The codification of ontological and contextual information in FunGramKB

The inclusion of ontological material in linguistic models is becoming increasingly popular. It implies the listing and description of basic concepts and their relations. In this section, the handling of contextual information and the inclusion of ontological material as in Mairal and Periñan-Pascual (2010) and Periñan-Pascual (2010, 2011, 2013), FunGramKB is discussed.

In my view, the two most important contributions of Faber, P.B and R. Mairal Uson (1999), and R. Mairal and P. Faber (2002) and in their extensive studies of the lexicon have to do

with the role they give to dictionary descriptions and the way they incorporate ontologies in their successive developments of their joint and /or separate works.

The inclusion of this material is analyzed in the work of Faber, P.B and R. Mairal Uson (1999), and R. Mairal and P. Faber (2002) More explicitly and with a direct computational orientation, Nieremburg and Raskin (2004) included an onomasticon among other improvements of their theory. Later on, an onomasticon also appears in Perriñan Pascual and Mairal Uson (2010).

The use of a dictionary definition as a starting point in linguistic modelization can be found in the work of Simon Dik (1989: 83), as what he calls “a stepwise lexical definition.” Such definition, he explains, is not very different from what a native speaker of a language and a monolingual dictionary would provide. Mairal Uson and P. Faber (2002: 40) also use this concept, which will be further developed in the design of FunGramKB and Corel as in Perriñan Pascual and Mairal (ibidem).

FunGramKB and its related conceptual representation language COREL, as in Perriñan-Pascual and Mairal Uson (2010) and Perriñan Pascual (2012) are based on the Lexical Constructional Mode. This, in turn, is presented in Ruiz de Mendoza and Mairal Uson (2008), in Perriñan-Pascual (2010), Perriñan-Pascual (2012), Mairal Uson (2002) and Mairal and Faber previous work give a base to the combination of ontological and lexicographic approach that led to their functional grammar knowledge base or FunGramKB. These developments form a cluster that has converged around FunGramKB and its related conceptual representation language or COREL.

Butler (2012) provides a thorough historical overview of these developments, along with his proposal of an ontological approach to the lexicon in a model called Functional Discourse Grammar.

In Perriñan Pascual (2012), the duplication of information in both the thematic frames and the meaning postulates is acknowledged, stating that thematic frames are fully integrated into meaning postulates. However, this duplicity is not solved

only by claiming that cognitive requirements meet the formalization of prototypical linguistic representations. Nor it is when referring to granularity problems.

In this same work Perriñan Pascual (ibidem) criticizes FrameNet, as in Fillmore (1982,1985) and Fillmore and Atkins (1992), and its successor Ruppenhofer et al. (2006), in terms that word senses are linked to manually inserted semantic frames to then automatically retrieve the syntactic patterns in which the lexical units are part of. However, apparently, this seems to be also the standard procedure used in most ontologies that use the lexicon of a language to derive meaning descriptions. In Perriñan Pascual (ibidem) critics, it is also regretted that the Frame Semantics model lies not only in the syntax-semantics interface but also in the deceptively (sic) deep approach to knowledge representation. It is argued there that the cognitive content of lexical units are described using a simple feature value matrix of conceptual relations, which is not enough. However, word senses in FrameNet lexical database are manually extracted and Framenet has had a long and successful life in NLP. Even in 2012, manual inclusion and / or extraction of data were not an unusual procedure.

The proposed alternative model is FunGramKB and Corel. The first is a Functional Grammar Knowledge Base and the second is a conceptual representation language. The ontology in FunGramKB is made up of Thematic Frames and meaning postulates. Both constructs seem to be fairly redundant because the first states the number and types of participants involved in a prototypical cognitive situation, portrayed by an event or quality, and the second focus more on the selection restrictions the two of them capture. Both thematic frames and meaning postulates involve different degrees of concretion /abstraction of the same basic concept. However, the vast majority of linguistic models recognize concepts such as predicate structure or logical form as basic underlying concepts, depending on the type and orientation of the linguistic formalism used.

It is also claimed there that these participants involved in such prototypical

cognitive situation portrayed in an event or quality, cannot always be instantiated linguistically, but that they are cognitively necessary. So if, as in Periñan Pascual (2012: 95), thematic frames are fully integrated into meaning postulates, this duplication of information would benefit from further explanations. It is also claimed that meaning postulates are used to store unsituated prototypical knowledge and that they are inherently intensional because they describe the mental representation of a feature that represents the applicability of the category of entities that the linguistic expression refers to. The point to be made here is that these underlying conceptual relations are codified in some languages, while in others they are not. If they were only cognitive, they must have been somehow present in most or all-human languages. But if they are language dependent, they cannot be part of a claimed language independent ontology. The reason is very simple because the first step in ontologies construction is supported by the semantic content portrayed in lexical items, the description being feature-based or not. Therefore, the position of language specific conceptual repositories, in relation with language independent ontologies, remains unclear.

Once again, it seems that the attempt to include a higher status concept to account for human cognition requirements is only obscuring the picture, rather than simplifying it. Again, cognition cannot be used as a linguistic specification without previously having delimited its different roles.

Also, in what is defined as a knowledge-engineering approach, an explanation of the mechanics of how the different parts of the Cognicon, Onomasticon, Thematic Frames, Meaning Postulates operate in proper computational terms would have been appreciated.

So, if, in a system claimed to be computational such as COREL/FunGramKB, this information is duplicated without its double formalization been justified, let alone explained, the problem of mixing the different functionalities of communication and cognition in a language application system, shows up.

Apart from this detailed and ontology based notational system, ways in which predications and meaning postulates can be computationally replicated would have been expected.

The provisional conclusion about the handling of ontological and contextual information in COREL/FunGramKB takes the form of a question that is how the duplication of information in predicate frames and meaning postulates is avoided.

Firstly, one is still left with the question of how different types of information, linguistic and nonlinguistic, are formalized in a way that a machine can understand them and, secondly, how these different types of information can be made compatible. Or, at least, how incompatibilities are explicit enough so that they can be further managed.

4.3. Other approaches to the codification of contextual and ontological information

Peñas, A. and E. Hovy (2010) propose a semantic enrichment of texts with background knowledge, and they claim to have attained some degree of synchronization between text representation, achieved by current NLP, and knowledge representation schemes that allow for further inference in text interpretation.

In their approach, background information is obtained in much the same way as in other NLP applications or indeed in FunGramKB. Making explicit the meanings attached to syntactic relations is a normal procedure in possibly all available linguistic models these days.

In this proposal, Peñas and Hovy (2010) show a clear explanation and representation of the selected identified concepts, and the procedures and forms for handling their data in a replicable way are clearly explained.

However, although the degree of granularity found in their BKB (Background Knowledge Base), their semantic descriptions are unexpectedly limited. Given the source of data Peñas and Hovy have used (a large corpus of American football descriptions), it may be used as a comparative reference since the database used in this paper is also a restricted type of lexicon,

such as the one used to describe wines. Because both are a restricted field, specific lexicons with common characteristics, Peñas and Hovy's analysis is of special value in the present case.

5. The DA as a procedure to handle different types of communicated information

The DA is not the first step of a new alternative linguistic model, and it does not claim to be so. It is just a way to handle linguistic and nonlinguistic related information for NLP purposes, clearly separating these and providing for ways in which redundant information can be handled. The main aim of the DA is to set the basis for further linguistic and nonlinguistic mark up.

The choice of an algorithm as a format is based on the claim Goded Rambaud (2007) of a structural similarity and compatibility of the grammar of a language, an ontology of an area of knowledge and the descriptive algorithm proper.

The grammar of a language is a simple construct. It only includes a series of identified grammatical entities as the particular language grammar categories that are relevant to the described language specifically and a series of language specific operational rules that link these identified categories. However, grammatical categories and grammar rules refer ultimately to the words that belong to a particular language.

As it is well known and extensively documented in typology studies, not all languages share the same types of grammatical categories. Even something as basic as parts of speech, although fairly general, is by no means universal. Let alone, grammar rules of particular languages.

Seemingly, ontologies, include only concepts. That is the conceptual entities recognized in an area of knowledge and the relations that link them. These are the basic rules that affect concepts in a general or in a particular area of knowledge. In contrast, with language dependent devices or word-based devices such as grammars, the basic units of ontologies are concepts. However, we humans have developed

words to name concepts and because of this there is a certain circularity when trying to define concepts separate from words.

On one hand, this distinction is important because there is often the case that the only way to capture a concept is to use the term that lexicalizes such concept. On the other hand, to identify a concept as part of the transmitted meaning in a language it must be linguistically described. As seen in previous sections, the formalization of a concept can take various forms, ranging from the linguistic description of a concept such as a dictionary definition, to complex ontological formula, or to just grammar notational variants, tailor-made to fit the parent linguistic model.

Since most common concepts are lexicalized, extracting their definition from a dictionary is a task that facilitates further identification of entities, relations, and common features by using simple parsers. Parsers identify class words first, and because usually nouns codify entities and verbs, adverbs and adjectives codify relations and types of relations, what a common parser does is ultimately provide a clue for ontological identification. The text that describes a concept in a monolingual dictionary can be used both as a text material for parsing and as a source for the ontological differentiation between entities, and the relations that link them.

In the example case of wine tasting descriptions, the basic concepts would include [FRUIT], [GRAPE], [FOOD], [DRINK] and [WINE], which are lexically described in glossaries. The identified relations include concepts such as [plant], [ripe], [prune], [harvest], [ferment], [age], [blend] and [bottle], all grammaticalized as English verbs, mostly codifying actions and processes.

Finally, an algorithm can be defined as a set of instructions to be performed in a certain sequence. An example of an algorithm could be something as simple as a cooking recipe: a list of components (ingredients) and a series of rules to handle them. That is, cooking actions to be sequentially performed.

Along the same lines, in Apresjan (2008: 53) it is claimed that every linguistic description

is ultimately made up of a grammar and a dictionary. This is not very far from what was originally proposed in the design of the descriptive algorithm in Goded Rambaud (2007), except the fact that in the DA, there is a separate slot for the dictionary definition and a general rule stating whether or not the sequentially rule is applicable in each particular case.

The DA includes some components that codify different kinds of information and a basic rule that handles linearity problems. Originally, it was just a minor scale attempt to account for the different types of information, both directly marked and ontologically preconfigured, which are relevant for the codification of the lexical material available in wine tasting notes.

In a precomputational stage, the DA agglutinates different types of information, ranging from nonlinguistic information, in the most to the left part of it; proper linguistic information, as part of speech and dictionary definition; to logical and ontological constraints, in the most to the right part of it. All these different types of information require some empirically documented codification.

$$\Sigma = [[(\alpha), (\psi), (\Delta), (\lambda), (\omega), (\mathfrak{D}), (\varepsilon), (\sigma), (\pi), (\theta), (\mu), (\delta), (\phi), (\beta)]. R]$$

There are several types of information that the descriptive algorithm captures. Firstly, nonlinguistic information and world knowledge that speakers, either share or retrieve from a commonly stored repository, is considered important. (α) ALFA codifies the most basic aspect of a term: the link to the perceptual input. For example, whether the lexical entry is or is not referential. For instance, the verb “stench” lexicalizes the experience of smelling repulsive odors, whereas there is no proper verb lexicalization of the experience of a delicious smell such as the smell of rose except saying that “it” smells like roses. The noun *aroma*, in both Spanish and English, codifies a pleasurable smell, but there is no symmetry in the type of lexicalization of negative or repulsive odours. One explanation could be that what both languages

There is usually phonetic, morphosyntactic, semantic and pragmatic markings which capture different types of meaning. However, in the case of cognitive components the evidence is based on the interpretation of the meaning of texts that the different authors provide. This is why up to now this type of information is given precise slots in the descriptive algorithm, in the hope to obtain some kind of specific linguistic codification of such cognitive constructs. Or at least, some other kind of codification or formalization of such meanings.

One such kind of capturing meaning using images is the use of emoticons in all kind of devices: iphones mainly. This involves some degree of abstraction and it is as old as ideographic writing, but in the present world where information of emotional states is achieved using images, iphone devices using emoticons prove that the hopes of the codification of cognitive meaning is not totally hopeless.

The original format of the DA as in Goded Rambaud (2007), including such types information, takes the following form,

mark is the non-pleasurable odor experience (Spanish: *apestar, heder*; English: *stink, stench*). So the pleasurable side of the smell experience is lexicalized in a more generic or abstract way, that is using the term *aroma* in both languages, while the disagreeable experience is lexicalized using verbs; that is involving an experiencer participant in the state of affairs. I claim that this is one more example of the use of language as a survival resource and its functionality as a communicative instrument. Because a bad smell characterizes a potentially dangerous state of food, the decomposition degree of food is a hint to assess how edible potential food still is. Consequently, it is important that such degree is lexicalized in a more detailed way. As a result, in this particular case, the degree of α in a scale will be high.

World Knowledge Information or Encyclopedic information is stored in (ψ)PSI. It is obviously possible to extract it from a multiplicity of on line resources, ranging from thesauruses to wikipedia or a simple Google search to be related to [Đ].

Next, specific linguistic information, in its most prototypical format, is captured in part of speech [λ], conventional parsers [ω] and dictionary entries [Đ]. Part of speech is captured in the λ slot with its subsequent notational variants such as POS (nouns λ 1), (verbs λ 2), (adjectives λ 3), (adverbs λ 4), and any parser usually does this job. Pragmatic & Discourse annotation that goes into omega (ω) and Dictionary entry (Đ) slots will be described below.

Following, non-proper linguistic information is the traditionally considered logical codification, where two basic types of logical structures are considered. They are the predicate structure and the qualia structure. Drawing from Aristotle, Katz 1972, proposes a model of the representation of entities as conceptual configurations different from predicates. In the DA, both are part of the logical construct [σ] and take the form of either a predicate [π] or the qualia [θ].

5.1. The component of dictionary definition

At this stage, where the lexicographical component becomes more relevant, some references are obligatory.

The component of dictionary definition Đ is pivotal in the design of the DA, and there is a tradition of noting this relevance with two different orientations. One group based in the USA and focused on the work of former Russian authors, such as Apresjan (2008) or Mel'čuk (1987) and another European group of authors, more focused on the theoretical explorations and connections that the linguistic analysis of a dictionary definition could provide.

Igor Mel'čuk's overall work in lexicography is by itself an essential reference. A very succinct resumé of his MTT (Meaning Text Theory) can be found as in Mel'čuk and A. Polguere (1987), where, in general terms, MTT is described as

a formalized model of natural language that simulates the linguistic behavior of humans.

A repository of lexicographic data about a particular language is stored in the so-called ECD (Explanatory Combinatorial Dictionary).

In Mel'čuk (1987: 262), it is explained how "(...) the ECD is consistently structured...this means that, with respect to its meaning and its cooccurrence, a lexeme is specified in terms of other lexemes". In addition, it is also explained that "(... a lexeme, which is fully determined in the lexicon by a hierarchical set of properties (...), one of this properties is the participation of the lexeme in question in the definitions of other lexemes in its own definition".

The most relevant characteristic of the definitions used in his Explanatory Combinatorial Dictionary (ECD) is that they are some paraphrasis or, as this author puts it, that they are "PARAPHRASE-BASED". Thus, the ECD as a database is, in fact, a formalized dictionary. Therefore, its possibilities of a direct integration of MTT in the DA, due to its most recent developments are obvious.

While Fillmore's influence in both WordNet, as in Fellbaum (1988) and the subsequent FrameNet is one that provides a higher degree of simplicity to the organization and shape of these databases, Mel'čuk's lexicographical contribution seems to provide a higher degree of descriptive detail to the grammar-morphosyntax-meaning complexity.

Within a functional tradition, the interest of the linguistic consequences and theoretical possibilities that the dictionary definition could eventually provide can be traced back to Dik (1989,1997) and Coseriu (1973,1977, 1978). It is in Dik (1989) where the concept of *definiens* is first given a lexicographical dimension further developed in Martin Mingorance (1990), in Mairal (2002). In Mairal and Faber (2002), this connection is also explored and the links between a dictionary definition and an ontology are first explicitly established.

In such a rich background and traditions, the DA includes the dictionary definition in two basic ways. Firstly, because as with conventional parsers, the automatic extraction of the Đ and

its insertion in the DA is fairly straightforward and had been done before. Secondly, and more important, because it constitutes a text that can be analyzed, parsed and directly linked to either any ontology used to feed the algorithm or to the specific ontological parts of the algorithm as such.

A dictionary definition can also be understood in terms of (conventional) meaning as in (Parikh (2007: 931) as “the semantic value of the form independent of any contextual factors” whereas “content is the contribution the form makes to the proposition expressed by the utterance within a sentential form in some context.”. This helps us focus on content as the object of analysis regarding the use of \mathcal{D} .

5.2. Logical, ontological and cognitive components

Logical and ontological components (ϵ), (σ), (π), (θ) are included in the DA because it seems that combined information from both linguistic and encyclopaedic sources only merges in ontologies. As it has been described above, the grammar of a language, the ontology of which the concept represented by the lexical entry under description is part of and the DA itself, are all three of them compatible conceptual structures.

As it was previously explained and as claimed in Nuyts (2004), communication and cognition are functionally different systems with different requirements. Cognition affects linguistic organization in the same way that it affects the handling of other types of information, apart from the specifically linguistic ones, and it also affects linguistic processing, though each one does so in a different way.

The most relevant contribution of cognitivism in linguistic description has to do with the idea that the way we perceive and process information coming from the world around us is determined by way we are physically shaped. It is somehow like saying that a computer cannot (yet) process and identify the different aromas and colors of wine because their software and

hardware do not allow them to do so..., yet. Cognition as such is reflected in language, but unlike logical and ontological concepts, it is less straightforwardly captured in linguistic terms.

Nevertheless, linguistic information mirrors certain cognitive processes which are currently modeled in the various existing cognitive linguistic models. Along these lines, the DA includes certain cognitive aspects that are not expected to be reflected in a specific type of conventional marking, but can be detected by statistical analysis. They are Lightly Codified Meaning (LCM) and they can be defined as transparent components, showing the fact their presence can be explicated and even described, but no particular linguistic marker identifies them in an exclusive way. Metaphors, for example, have to be identified using procedures other than looking for conventional phonetic, morphosyntactic or lexical markers. For example, using corpus statistical analysis as in Goded Rambaud (2007).

Because the main objective of the descriptive algorithm is to set the basis for further proper computational tagging, the ontological component has been given a relevant role in the DA, where it provides for two ontological approaches simultaneously. One of them is the traditional ontological/logical division between relations and entities as in Lyons (1977: 325) and another is a cognitive configuration of ontological material. These two approaches sometimes complement each other, but occasionally they produce overlapping information.

The available scope of both ontologies and models of grammars is wide. Some of them are tailor made to suit certain linguistic paradigms, and others are also tailor-made to suit certain ontological requirements for NLP.

Therefore, the ontological and cognitive configuration that allows language users shape other types of information is captured in the DA. Some ontological material is deeply rooted in the stable properties of our empirical experience of the world, or the communication of concepts

would not be possible. Nevertheless, there is also another part of ontological material that can be affected by the cognitive configuration of the world that humans have. This fact has two consequences. One is the need to consider different possibilities of available formats for ontologies and assess the way they account for other than non-conventional linguistic information. The other is the need to reorganize the configuration of the ontological material in the DA.

While humans draw nonlinguistic information from a variety of sources in addition from the linguistically codified one, for a machine to be able to recognize a linguistic pattern, it must be lexically, morphosyntactically or textually recognizable, which takes us back to their linguist modelization looking for further types of tagging. In addition, machines are increasingly more able to recognize image patterns. It is not far away the day when they will be able to recognize other sensory perceptions too.

What is not that easy, as if it were ever easy in the first place, is the modelization of cognitive material. For example, Perrián-Pascual (2010), and Perrián and Mairal (2010) propose the storage of procedural knowledge such as scripts in their Cognicon but no formalization of how this can be done is yet offered. It could be guessed that they refer to different linguistic versions of scripts.

In Paradis (2008) certain cognitive-ontological aspects described there can be identified in certain collocations. Because of this, such aspects are included in the DA. Its seems that if certain features of meaning can adopt certain collocational patterns which can be identified in texts they should be interpreted as part of ontological markers. For example, in Paradis (2008: 317-343), in the following example:

a very good book, disgusting, totally agree, what a car, thoroughly enjoy, very British, a very key strategic question, pretty crap

a configurational meaning can be identified. In the DA, such configurational meaning is marked

using the convention $[\varphi]$. Humans can identify a degree pattern and/ or a simple quantifier, but a machine needs further cognitive-ontological mark up to identify them. The next problem is how.

One possibility, involving circularity, is to search a corpus for certain expressions, previously identified, as linked to cognitive interpretations and expand this interpretation to the recurrent collocations marked accordingly.

As a result, the configuration of the ontological material in the DA is provided in $[\epsilon]$ Type of ontological category, $[\mu]$ part/whole schematic ontology, $[\delta]$ Degree, $[\varphi]$ and $[\beta]$ Boundedness. Clearly, some overlapping in the marking of the type of ontological and logical categories is expected. Only the application of the linearity feature will force to select one or the other. Since both ontologies and logical representations are hierarchical, the backwards effects of R will make such DA component to be taken into account or not.

The Thing/Relation option is considered relevant at the higher level of abstraction. It is subsumed in (ϵ) and thus, not included in a schematic ontology. Other Paradis' schematic ontologies such as focus, order, and modality again are considered at different levels of Concretion/Abstraction and not developed in the present development of the DA.

The clashing of information among components is frequent, and it is used as a procedure to identify semantic incompatibilities in the use of terms, such as *decadent*, as in Goded Rambaud (2007). This concep, similar to what had been termed in the Chomskian tradition as selection restrictions, is a procedure for confronting information from the different components in the DA and it was labelled as clashing identification procedure (CIP). It's based on the confrontation of statistic collocational information that helps us identify when certain adjectives are used metaphorically or referentially. See, for example, the collocational pattern of *decadent* in both Robert Parker's and the BNC corpus:

Wine Tasting Corpus (WTC)				British Nacional Corpus (BNC)					
Collocates				Collocates					
	Total	CW (3/4)	Total WTC	as C (3/4)		Total	CW (3/4)	Total BNC	as C (3/4)
de	86	21,2	81039	0,1	sensuality	2	2,06	148	1,37
de	69	14,7	17946	0,33	thrill	2	2,06	482	0,41
de	46	11,6	40877	0,11	old	1	1,04	62832	0

FIGURA 1

The case o a decadent wine.
The D component

Thus statistic collocational information of the term *decadent* was extracted from at least two corpora, as in Goded (2010). This evidence was used to identify referential and non-referential uses, by means of a Clashing Identification Procedure (CIP) and thus, to help disambiguation. For example, the term *decadent* collocates differently depending on the corpus. If in the BNC it nearly ever collocates with *wine* and in the WTC it collocates with *wine*, *rich* and *fruit*, it can be said to have been used metaphorically in the WTC and referentially in the BNC.

Therefore, this simple system of amalgamating different types of information and comparing collocational statistic empirical data adds support to the validity of using the descriptive algorithm as one possible procedure, at a precomputational stage, for further NLP uses. For example to help text disambiguation.

5.3. Linearity or sequentiality in the DA

The idea of agglutinating different types of information in a single representational tool arises from the fact that the human mind processes simultaneously not only linguistic information but also other types of information together with the linguistic one. The multimodality of the sensory experience is particularly interesting in the case of wine tasting and the fact that these experiences are nearly simultaneously perceived, contrasts with the fact that the linguistic expression used to describe it is necessarily sequential. And it is sequential, not because of the textual particularities of a wine tasting note but because human language is linear. The fact that language is linear was first put forward by Saussure (1945) as a key characteristic of language.

Thus, linearity or sequentiality is the general rule of organization of the components in the DA and it is represented as R. It determines whether a certain component is or is not applicable depending on its particular nature. For example, sequentiality affects logical, grammatical and lexical linguistic codification but it does not affect ontological components

as such. Ontologies are made up of concepts and, although they are naturally hierarchical, such hierarchy affects the sequence of their relations but not the internal configuration of the concepts per se.

To sum up, the DA amalgamates types of information that are captured in components of different type, quality, and scope of influence. Not all these components are of purely linguistic nature. They range from prototypically nonlinguistic information to genuinely linguistic and ontological one, highlighting those features that are more relevant in the case of particular semantic fields, corpus or areas of knowledge, such as in the case of wine tasting expressions.

It is claimed in Goded Rambaud (2007, 2010, 2012) that, because of the language property of being linear, whether sequentiality is or is not applied to the components of the Descriptive Algorithm, is a factor to be considered because it can disregard much information as redundant or irrelevant.

The issue of the sequentiality or linearity property of language is also related to semantic compatibility because it only applies to syntagmatic relations.

The components of the DA can be incompatible in many ways and at different levels of abstraction, but the simple fact that incompatibility among them can be identified provides negative evidence well worth exploiting. This is why the identification and description of a more cognitive than linguistic phenomena such as metaphorization or intersubjectivity should be separated from the way these phenomena are linguistically either marked or left unmarked.

6. Case studies

6.1. Case study 1. Wine metaphors. The case of “elegant” and “decadent” wines

Partly because of the generalization of an interest in wine in general and partly because of its sociological prestige, wine tasting description is receiving increased attention in linguistics. Following the initial work of Lehrer’s pioneering

linguistic description of wines (1975, 1983), other linguistic studies include Lehrer (1992, 2002), Caballero (2007, 2012), Goded Rambaud (2008) and Paradis (2005, 2009, 2010, 2011, 2002, 2013) and Paradis and Kerren *et al.* (2011).

Metaphorization in wine tasting notes has been well described in Paradis (2010, 2011, 2013); Caballero (2007) and Caballero and Paradis (2015), Suárez Toste (2007) and others. A discourse analysis of wine description is found in Hommerberg (2011). And the use of winespeak as a database extracted from Robert Parker's Wine Advocate 2007 can be found in Goded Rambaud (2007, 2010, 2012).

The linguistic description of a sensory experience can be prototypically found in the form of wine tasting notes (WTN). In its prototypical form, this is a type of text that describes the sequence of seeing, smelling and tasting the wine in this precise order. This type of lexicalization necessarily encapsulates a multimodal, nearly simultaneous, sensory experience into the linear and sequential constrictions of linguistic codification.

The lexical entries *wine*, *caress*, *elegant*, and *decadent* are first discussed as cases of wine metaphors whose interpretation as such need other than conventional linguistic codification for NLP purposes.

Goded Rambaud and Ibáñez (2012) used Mairal and Faber's (2002) lexical templates and their developments in Mairal and Ruiz de Mendoza and Mairal Uson (2008) and in Perrián Pascual and Mairal (2010) and in Perrián Pascual (2013) to develop the lexical entries, *wine*, *caress* and *elegant*. However, apart from the problem of total manual tagging, it was impossible to replicate the procedure because the mechanics of the application were not available then.

The term 'caress' was first discussed in Goded Rambaud and Jimenez Briones (2002). Later on, this term, as in Robert Parker (2007) The Wine Advocate, was extracted as follows:

(...) *a wine that caresses the palate with gentle tannins and sweet, creamy oak.*)

Ref.: RP 2002 'G' Pago la Jara.

The verb 'caress' had been previously explored by Goded Rambaud and Jiménez (2002: 323) to identify how the then called lexical templates could integrate Mairal's Functional Lexematic Model (2002: 39) and Van Valin's Role and Reference Grammar. Goded Rambaud showed how *touch* and *hit* could be differentiated by the so called intensity parameter. In between the touch→hit continuum, this parameter can be used to mark the lexicalization of *touch* hyponyms such as *tickle*, *pat* or *stroke*. At the other end of the continuum, it can be used to show how it works in verbs like *strike*, *bash*, *punch*, *slap*, *beat*.

Later on, in Goded Rambaud and Ibáñez (2012), it is discussed how the representation of CARESS shows in thematic frames and meaning postulates in FunGramKB and Corel. The description using such modellization is as follows:

(...) CARESS: [do'(w, [use.hands.fingers. (x) a.pleasant.manner' (w,x)])] CAUSE [[do' (x, [touch'(x,y)])]] $\alpha = x$ (touch' [MANNER 1 softly, 2 slowly, 3 carefully] (x, y))

Where the argument 'w', an instrument that refers to the quality of the wine and its texture, could be cognitively interpreted as a metaphor because of what the wine does to the palate. It is then a comparison cognitive tool. In this particular case, the predicate schema could be used as a form of lexical representation. In another example, the text is extracted from Robert Parker's 2007 Wine Expectator¹, which is used as the source of data.

WE356 Defines Tchelistcheff's,

...as iron-fist-in-a-velvet-glove description of Stags Leap District Cabernets. It is Seductively soft, a wine that **caresses the palate** with gentle tannins and sweet, creamy oak, framing ripe blackcurrant fruit and rich spice. The power is in the focus, brilliance, and overall purity.

Thus, the terminal concept CARESS (ref. WE356) would be represented using COREL as:

(\$CARESS_00).

According to with such system of conceptual representation, it will be configured as an event, and it will be subordinate to basic event

– +TOUCH_00.

It is also possible to build the *caress* template with all adverbial modifications needed for manner. In this case, *touch* is referentially used. However, this lexical entry is turned metaphoric when it collocates with other touching sensory descriptions.

The metaconcept derives the perception thematic frame in:

(...) #PERCEPTION: (x1) Theme (x2) Referent

Theme: Entity that perceives another entity through any of the senses.

Referent: Entity that is perceived through any of the senses. This shows how in that model *caress* inherits this scheme and from there, its meaning postulates add the *caress* specifications in the form of satellites and operators. But this inheritance factor claimed by COREL and FunGramKB creators is something that any simple online ontology will do without this complex apparatus and the heavy load of manually marking. Besides, possible redundancy problems have not even acknowledged.

Exploring further the Longman Exams Dictionary (p. 209), the following definitions for *caress* are found.

1. 'to touch someone gently in a way that shows you love them'.
2. 'to touch something gently, in a way that seems pleasant or romantic'.

And the representation of *caress*, using only meaning postulates as in Goded Rambaud and Ibáñez Moreno (2012: 181-196) shows as follows:

(..) Meaning POSTULAE of *caress*:

+ (e1: +TOUCH_00 (x1: +HUMAN_00) Theme (x2) Referent (f1: +GEN-TLE_00) Manner & (f2: (e2: +LOVE_00 (x2) Agent (x1) Theme)) Reason) * (e2: +TOUCH_00 (x3: BEING_00) Theme (x4) Referent (f3: +PLEASANT_00 ^ +ROMANTIC_00) Manner).

However, no empirically identifiable data, either lexical, grammatical or textual of

how, other than manually, this assignment of features can be achieved. It seems that it is only the intuition of the human tagger and his/her tagging work what is apparently used. Besides, the procedure of marking all this prototypical ontological information remains unclear. If no specific linguistic marking is suggested for English or any other language, it seems that a manual marking of each and all identified prototypical features in each case is the obvious answer. If this is so, doubts can be raised about how this approach handles so-called engineered knowledge models.

This type of problems found in models claiming to capture ontologies and world knowledge information possibly suggests more modest approaches, trying to use what is presently available as ontological and knowledge resources. The combination of Dictionary definition extraction followed by subsequent parsing of such definition together with a comparative statistical analysis of collocations offers interesting perspectives.

Using the descriptive algorithm, the dictionary definition of *caress* presents information encapsulated in [D] and it can be directly imported from any online dictionary rather than manually incorporated. Two examples are offered to discuss the role of the dictionary definition in relation with ontologies for the terms *wine* and *elegant*.

As explained above, the DA includes a central component, the dictionary definition, around which all other components are included. Particularly, the position of one word in ontological terms can be deduced from its POS tagging and from the parsing of the text extracted from the dictionary. On one hand, the use of the term *wine* in its conventional and referential dictionary definition as the [D] is studied, and a comparison of ontologically oriented description can be offered. The dictionary entry for *wine* was taken from the Collins Cobuild (Sinclair (1987: 1673)) is defined as,

(..) "An alcoholic drink which is made from grapes and is usually red or white" which can be rewritten regarding an hyponymic relation (A KIND OF) as,

(.) A wine is a kind of liquid, and it can be represented as encapsulating the following types of information,

{[(λ),(ω), (Ð),(ε1),(θ),(δ),(β)].R}

In an example extracted from The Wine Advocate 2007 (ref. WE356), the development of the component (ω) explores the example *caress* to show the possibilities of its ontological extraction from dictionary definition [Ð] in the DA. On the other hand, the use of the DA for wine as a drink includes the following descriptors,

(.) wine

{[(α) (λ:noun), (σ1), (θ),].R}

These formalisms are now compared with the DA. If we do wish to represent the term *wine*, using the descriptive algorithm, it shows as follows,

wine = {[(α), (ψ), (Δ), λ:noun), (σ1), (θ),].R}

- (α) ALFA codifies the most basic aspect of a wine tasting note: the link to the perceptual input. That is, whether the lexical entry has or has not a connection with human sensory sources of information. For example, the verb *stench/ apearstar*, in both English and Spanish, lexicalizes repulsiveness, whereas there is no lexicalization of the smell of a rose, except saying that “it” smells like roses. The word *aroma*, the same term in both Spanish and English, codifies a pleasurable smell. Both the noun *aroma* and the verb *stench / apearstar* are directly connected with the human sensory perception systems. Any English dictionary definition of such verbs will require some reference to the human sense of smell.
- (ψ); PSI codifies World Knowledge Information or encyclopedic information. The obvious connection is Wikipedia. It is indeed necessary to differentiate this beverage from others.
- (Δ), DELTA codifies the link to the referent. It just marks whether the word is used

referentially or not and, if not, therefore triggering or hindering metaphorical readings.

- (Ð)] capital D, codifies dictionary definition. This is one of the most important aspects of the descriptive algorithm because it can be automatically extracted from any monolingual dictionary and this text subsequently parsed.

Logical and ontological markers include (ε) EPSILON that codifies the type of logical category in terms of order of entity. (σ) SIGMA that codifies the type of ontological construct in terms of whether it is an entity σ1 or a relation σ2. In this case, wine is an entity. (θ) ZEDA codifies entities as logical constructs.

No further cognitive markers are applicable to this term in isolation unless we are provided with collocational evidence showing that a certain expression codifies certain cognitive operation or construct.

Additionally, as in Goded 2010, a simple ontological approach in the case of a *decadent* wine shows how a clashing underlies metaphorization in the ontological configuration of the properties of the adjective *decadent*. It also shows that the relation between the default referent for *decadent* in the WTC (Wine Tasting Corpus) and its prototypical referent, can be highlighted comparing the ontological status of both using the (SUMO) ontology:

A decadent1 person {person: Cognitive agent < Agent < Object < Physical entity < Entity}

A decadent2 wine {wine: Beverage < Food < Object < < Physical entity < Entity}

Here the metaphor is not only changing the semantics of *decadent* but modifying the concept of wine altogether and making it more flexible. Furthermore, a comparative statistical analysis of both Robert Parker 2007 and the BNC of these terms collocations as in Goded Rambaud (2010) shows that if there is a clashing among DA components, this fact could be used as a possible metaphor identification utility. It is claimed in Goded (ibidem) that an ontological

analysis and classification of clashing using the comparison between the more referential and the more metaphoric collocations, depending on its context (the corpora they have been in) is a simple procedure that could facilitate machine identification of non-linguistically codified and cognitively preconfigured lexical elements.

7. Conclusions

One of the main problems a text presents at its precomputational stage has to do with the selection of concepts that produce (or not) proper empirical evidence. Be it phonetic, morphosyntactic, semantic, lexical and grammatical evidence in a certain language. This, in turn, has to do with the preferred model or group of models that claim to shape such selection. Although it seems that cognition and language as human capacities usually interact, the linguistic models claiming to incorporate cognitive information in their design fail to show how this information can be empirically identified at the different levels and in the systems that human languages codify them. The problem is that sometimes the way in which human cognition affects linguistic codification is not marked in a clearly identifiable way. Because of this, LCM (Lightly Codified Meanings) identified in Goded Rambaud (2007, 2009, 2012) are proposed as a tool to handle important identified areas of meaning such as intersubjectivity or the wording of the wine tasting experience.

Part of the problem when trying to codify ontologic and contextual information focuses on a lack of differentiation between the communicative and cognitive functions of language in the design of certain linguistic models. Because methodological differences between explanation and description affect the way such meanings are marked in particular linguistic models, it is claimed here that not differentiating between these functionalities affects both linguistic description and linguistic modelization.

Instead of using whole self-contained linguistic models for precomputational codification purposes, a simple descriptive

algorithm is used as an easier device for extracting certain types of information from ontologies and dictionaries in order to precodify unclear meanings. This DA is a simple formalism that encapsulates related information for NLP purposes. The DA intends to account for linguistic and nonlinguistic information relevant to the description lightly codified meanings capturing sensory experiences such as, for example, the wine tasting experience.

The choice of an algorithm as a blueprint to capture the various types of information transmitted in communication exchanges is based on the Goded Rambaud (2007) claim of a structural similarity, and compatibility of the grammar of a language, an ontology of an area of knowledge and the proper descriptive algorithm. The grammar of a language is a simple construct. It only includes a series of identified grammatical entities as categories relevant to the described language and a series of language specific operational rules that link these identified categories. As it is well known and extensively documented in typology studies, not all languages share the same types of grammatical categories. Even something as basic as parts of speech, although fairly general, is by no means universal. Let alone, grammar rules of particular languages. However, grammatical categories and grammar rules refer ultimately to the words that belong to a particular language.

Along the same lines, ontologies include the entities recognized in an area of knowledge and the basic relations affecting the linking of such entities. In contrast with language dependent devices or word-based devices, such as the dictionaries or the grammars of specific languages, the basic units of ontologies are not words but concepts. Finally, an algorithm can be defined as a set of instructions to be performed over a set of entities in a certain sequence. As a result, these three constructs have important similarities.

The descriptive algorithm (DA) includes contextual information, referential information, types of linguistically codified information, the dictionary definition of the lexical entry under analysis, logical and ontological information, and

eventually and, only if this particular cognitive information has been lexically and/or textually identified in a corpus in collocational terms, cognitive information as well. Since sequentiality is a key linguistic factor in linguistic codification, the components of the algorithm may or may not be affected by it.

The DA includes the dictionary definition for two basic reasons. Firstly, because automatic extraction of dictionary definition and its insertion in the DA is a fairly straightforward procedure. Secondly, because it constitutes a text that can be analyzed, parsed and directly linked to either an ontology used to feed the algorithm or to the specific ontological parts of the algorithm as such. This device was originally designed to agglutinate information at a precomputational stage in a Wine Tasting Notes (WTN) corpus because it captures the multimodality of this type of sensory descriptions. Since WTN are pieces of language heavily affected by the context that produces important effects in the constraints of their linguistic codification, they are also used as examples of LCM (Lightly Codified Meanings).

It is claimed here that, given the fact of the lack of a unified linguistic model, based on an agreed and shared body of basics that capture specific ontolinguistic and contextual codification restrictions on top of the traditional phonetic, morphosyntactic and semantic-pragmatic ones, simpler devices for the organization of linguistic and nonlinguistic information may be more practical for a pre computational treatment of natural languages. Thus, the descriptive algorithm is proposed, not as another alternative linguistic model, but as an alternative way of capturing the ontological, contextual and linguistic information transmitted among humans in a type of format that is expected to facilitate its further computational processing.

And finally, one problem of most linguistic models claiming to account for different types of information is that they require an important amount of manual tagging because, up to date, only humans are able to identify certain types of cognitive and contextual information. This manual tagging load can be reduced if, instead of marking each lexical item for different levels

and types of information, the human tagger only has to mark bunches of information for one single predefined component. As a result, filling in the DA slots could be an easier and less time-consuming alternative to manual tagging.

Note

1. Robert Parker 2007 database, as an Access file, including references from the *The Wine Advocate* was used by Margarita Goded under authorization. It is a corpus of wine reviews taken from the American wine magazine *Wine Advocate*, under Parker's direction. This database also includes references from Parker's wine tasting notes published at *The Wine Expectator* review. It can be found at: <http://www.winespectator.com/>

Referencias/ References

- Apresjan, J. D. 2008. "Principles of Systematic Lexicography". In: Thierry Fontenelle (Ed). *Practical Lexicography. A Reader*. United Kingdom: Oxford University Press.
- Butler, C. 2012. "An ontological approach to the representational lexicon in Functional Discourse Grammar". In: *Language Sciences XXXIV* (5): 619-634.
- Butler, C. S. and J. Arista, Eds. 2009. *Deconstructing constructions* (Vol. 107). John Benjamins Publishing.
- Caballero, R. 2007. "Manner of motion verbs in wine description". In: *Journal of Pragmatics XXXIX*: 2095-2114.
- Caballero, R. and C. Paradis. 2015. "Making sense of sensory perceptions across languages and cultures". In: *Functions of Language*, XXII (1): 1-19.
- Clark, R. and P Parikh. 2007. "Game theory and discourse anaphora". In: *Journal of Logic Language and Information*. (2007) XVI: 265. doi:10.1007/s10849-006-9037-7

- Connolly, J.H. 2007. "Context in functional discourse grammar". In: *Alfa*. Sao Paulo. LI (2): 11-33.
- Connolly, J.H. 2011. "The Contextual Component within a dynamic implementation of the FDG model". Paper presented at the Workshop on the Contextual Component in Functional Discourse Grammar. University of Barcelona, Spain, 8-9-September, 2011.
- Coseriu, E. 1973. *Teoría del lenguaje y lingüística general* (Vol. 1). Madrid: Gredos.
- Coseriu, E. and M. M. Hernández. 1977. *Principios de semántica estructural*.(Vol. 259). Madrid: Gredos.
- Coseriu, E. 1978. *Gramática, semántica, universales: estudios de lingüística funcional*. Madrid: Gredos.
- Cruse, D. A. (Ed.). 2002. *Lexicology: an international handbook on the nature and structure of words and vocabularies*. (Vol. 21). Walter de Gruyter.
- Dik, S. 1989. *The Theory of Functional Grammar. Part I: The Structure of the Clause*. Foris Publications. Dordrecht-Holland/Providence RI- USA.
- Dik, S. C. 1997. *The theory of Functional Grammar: Part 2: Complex and derived constructions*. Hrsg. von Kees Hengeveld. Berlin: De Gruyter Mouton.
- Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Fontenelle, T. 2008. *Practical lexicography: a reader*. Oxford: Oxford University Press.
- Goded Rambaud, M. and R. Jimenez Briones. 2002. "Lexical-semantic explorations in English verbs of physical contact: Iconicity and linguistic representations". In Mairal, R and M.J. Pérez-Quintero (eds). *New Perspectives in argument structure in Functional Grammar*. Berlin/New York: Mouton de Gruyter. Pags.: 303-327.
- Goded Rambaud, M. 2007. "Un algoritmo descriptivo para un corpus léxico sobre cata de vinos". In: *Actas de los xii Encuentros Internacionales sobre Sistemas de Información y Documentación* (IBERSID 2007): 313-321.
- Goded Rambaud, M. 2008. *La percepción sensorial y el léxico de la comida y la bebida en inglés y en español*. Festschrift volume in honor of Asunción Alba Pelayo. Madrid: UNED University Press.
- Goded Rambaud, M. 2009. "A descriptive algorithm for a wine tasting lexicon corpus". In: *Scire: Journal of information and documentation systems*. XV (2): 39-62.
- Goded Rambaud, M. and A. Poves. 2010. (Eds) *Proceedings of the First International Workshop on Linguistic Approaches to Food and Wine Description: 57-72*. Madrid: UNED.
- Goded Rambaud, M. 2010. "Can taggers and disambiguators be theory free?: the search for a unified approach in lexical representation". In: *Modos y formas de la comunicación humana* (pags. 1091-1102). Servicio de Publicaciones. Universidad de Castilla La Mancha.
- Goded Rambaud, M. 2012. "Intersubjectivity and persuasion in wine tasting notes". Paper presented at the Workshop on Modality, Corpus, Discourse at the University of Lund. June 2012.
- Goded Rambaud, M. and A. Ibáñez Moreno. 2012. "El desarrollo de un etiquetador semántico-cognitivo para el procesamiento

- de las entradas léxicas de un corpus de notas de cata de vino”. In: *Vino y alimentación: estudios humanísticos y científicos*. Universidad de La Rioja: pags.: 181-196.
- Goded Rambaud, M. 2015. *The Codification of Meaning in English*. Madrid: McGrawHill.
- Goded Rambaud, M., A. Ibáñez, and V. Hoste. 2015. *On the Architecture of Words. Applications of Meaning Studies*. Madrid: UNED University Press.
- Grisham, R. 1986. *Computational Linguistics. An Introduction*. GB: Cambridge University Press.
- Hengeveld, K., and L. Mackencie. 2008. *Functional Discourse Grammar: A Typologically Bases Theory of Language Structure*. Oxford and New York: OUP.
- Hommerberg, C. 2011. *Persuasiveness in the discourse of wine: The rhetoric of Robert Parker*. Doctoral Thesis. Växjö, Kalmar, Linnaeus University Press. 262.
- Hymes, D. 1972. On communicative competence. *Sociolinguistics*: 269-293.
- Katz, J. J. 1972. *Semantic theory*. New York: Harper & Row.
- Kerren, A. et al. (2011, July). “Visualization of sensory perception descriptions”. In *2011 15th International Conference on Information Visualisation*: 135-144. IEEE.
- Kreidler, C. W. 1998. *Introducing English semantics*. Psychology Press.
- Langacker, R. W. 1987. *Foundations of Cognitive Grammar. Vol. I Theoretical Prerequisites*. Stanford, California: Stanford University Press.
- Langacker, R. W. 1991. *Foundations of Cognitive Grammar. Vol. II Descriptive Application*. Stanford, California: Stanford University Press.
- Lehrer, A. 2002. “Paradigmatic relations of exclusion and opposition I: Gradable antonymy and complementarity”. In Cruse, D. A. (Ed.). (2002). *Lexicology: an international handbook on the nature and structure of words and vocabularies*. XXI: 498-507. Walter de Gruyter.
- Lehrer, A. 1975. “Talking about wine”. *Language* 51(4): 901-923.
- Lehrer, A. J. 1983. *Wine and conversation*. Bloomington: Indiana University Press.
- Lehrer, A. J. 1985. “Is semantics perception-driven or network-driven?” In: *Australian Journal of Linguistics*. V (2): 197-209.
- Lehrer, A. and E. F. Kittay. 1992 *Frames, Fields and Contrasts*. Hillsdale, New Jersey: Lawrence Earlbaum Associates, Publishers.
- Lyons, J. 1995. *Linguistic Semantics. An Introduction*. Cambridge USA. Melbourne Australia: Cambridge University Press.
- Mairal Usón, R. y P. Faber. 2002. “Functional Grammar and Lexical Templates”. In: R. Mairal y M.J. Pérez Quintero (eds.) *New perspectives on predicate argument structure in Functional Grammar*. Berlin and New York: Mouton de Gruyter: 41-98.

- Mairal Usón, R. y F. Ruiz de Mendoza. 2008. "Levels of description and explanation in meaning construction". In: Ch. Butler y J. Martín Arista (eds.) *Deconstructing Constructions*. Ámsterdam/ Philadelphia: John Benjamins: 153-198.
- Mairal, R and M. J. Pérez-Quintero (ed.). 2002. *New Perspectives in argument structure in Functional Grammar*. Berlin/New York: Mouton de Gruyter.
- Mel'čuk, I. and A. Polguere. 1987. "A Formal Lexicon in the Meaning Text Theory: (Or How to Do Lexica with Words). In: *Computational Linguistics XIII* (3-4): 261-275 .MIT Press Cambridge, MA, USA
- Mitkov, R. 2003. *The Oxford Handbook of Computational Linguistics*. Great Britain: Oxford University Press.
- Niremburg, S. and V. Raskin. 2004. *Ontological Semantics*. Cambridge, Massachusetts. London, England: MIT Press.
- Nuyts, J. 2001a. *Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective* (Vol. 5). Amsterdam: John Benjamins Publishing.
- Nuyts, J. 2001b. "Subjectivity as an evidential dimension in epistemic modal expressions". In: *Journal of Pragmatics XXX*: 383-400.
- Nuyts, J. 2004. "Remarks on layering in a cognitive-functional language production model". In: *FUNCTIONAL GRAMMAR SERIES*: 275-298.
- Nuyts, J. 2007. "Cognitive linguistics and functional linguistics". In: D.Geeraerts, H.Cuyckens (eds.), *Handbook of cognitive linguistics*. Oxford: Oxford University Press: 543-565.
- Nuyts, J. 2012. "Notions of (inter)subjectivity". In: *English Text Construction V*: 53-76.
- Parker, R. See *The Wine Advocate /The Wine Expectator*.
- Paradis, C. 2005. "Ontologies and construals in lexical semantics". In: *Axiomathes XV* (4): 541-573.
- Paradis, C. 2008. "Configurations, construals and change: expressions of DEGREE". In: *English Language and Linguistics XII* (02): 317-343.
- Paradis, C. 2009. "This beauty should drink well for 10–12 years: a note on recommendations as semantic middles". In: *Text & Talk-An Interdisciplinary Journal of Language, Discourse Communication Studies XXIX*: 53-73. Berlin: Mouton de Gruyter.
- Paradis, C. 2010. "Touch downs in winespeak: ontologies and construals in use and meaning making". In: Goded and Poves (eds) *Proceedings of the First International Workshop on Linguistic Approaches to Food and Wine Description*: 57-72. Madrid: UNED.
- Paradis C. 2011, "Visualization of sensory perception descriptions". In: *2011 15th International Conference on Information Visualisation*. IEEE.. In Kerren, Andreas, 2011. "Visualization of Sensory Perception Descriptions Proceedings". 15th International Conference Information Visualisation. Editorial: IEEE: 135-144.

- Paradis, Carita and Caroline Willners. 2011. "Antonymy: From convention to meaning-making." In: *Review of cognitive linguistics* IX(2): 367-391.
- Paradis, C. and M. Eeg-Olofsson. 2013. "Describing sensory experience: The genre of wine reviews". In: *Metaphor and Symbol* XXVIII (1): 22-40.
- Parker, R. (2008). The Wine Advocate Rating System. Obtained from: <https://www.erobertparker.com/sitesearch/rparker.aspx>
- Parikh, P. 2007 "Game theory and discourse anaphora". In: Clark, R. and P. Parikh. *Journal of Logic, Language and Information* XVI (3): 265-282.
- Parikh, P. 2007. "Situations, rules, and conventional meaning: Some uses of games of partial information". In: *Journal of Pragmatics* XXXIX (5): 917-933.
- Parikh, P. 2010. *Language and equilibrium*. Cambridge, Massachussets: The MIT Press.
- Parker, R. 2008 *Parker's Wine Buyer's Guide* (7th ed.). Simon & Schuster. ISBN 978-0-7432-7199-8.
- Peñas, A., and E. Hovy. 2010. "Semantic enrichment of text with background knowledge". In: *Proceedings of the NAACL HLT 2010 First International Workshop on Formalisms and Methodology for Learning by Reading*: 15-23. Association for Computational Linguistics.
- Periñan Pascual, C. 2012. "The situated common-sense knowledge in FunGramKB". In: *Review of Cognitive Linguistics* X (1): 184-214.
- Periñan Pascual, C. 2013. "A knowledge-engineering approach to the cognitive categorization of lexical meaning". *VIAL: Vigo International Journal of Applied Linguistics* X: 85-104.
- Periñan Pascual, C. and R. Mairal .2010. "La gramática de COREL: un lenguaje de representación conceptual". In: *Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile* (21): 11-45.
- Ruiz de Mendoza Ibáñez, F. J. and R. Mairal Usón. 2008. Levels of description and constraining factors in meaning construction: an introduction to the Lexical Constructional Model. *Folia Linguistica*.
- Saeed, J. 2016. *Semantics*. 4th ed. Singapore: WILEY Blackwell.
- Schalley, Andrea and Dietmar Zaefferer eds. 2007. *Ontolinguistics. How Ontological Status Shapes the Linguistic Coding of Concepts*. Berlin: Mouton deGruyter.
- Saussure, F. de. 1945. *Curso de Lingüística General*. Buenos Aires: Editorial Losada.
- Sinclair, J. Ed. 1987. COLLINS COBUILD ENGLISH LANGUAGE DICTIONARY. London. Glasgow: Harper Collins Publishers.
- Suárez Toste, E. 2007. "Metaphor inside the wine cellar: On the ubiquity of personification schemas in winespeak". In *Metaphorik de* XII (1): 53-64.
- Fontenelle, T.(Ed.). 2009. *Practical Lexicography. A Reader*. United Kingdom: Oxford University Press.

Van Valin, R. D. and R. J. LaPolla. 1997.
Syntax: Structure, meaning, and function.
Cambridge University Press.

Vossen, P. 2003. *Ontologies.* In Ruslan Mitkov ed
*The Oxford Handbook of Computational
Linguistics.* Great Britain: Oxford
University Press.

Wine Spectator Press, New York, NY (1999).

Dictionaries used

COLLINS COBUILD ENGLISH
LANGUAGE DICTIONARY. Glasgow:
HarperCollinsManufacturing. ISBN 0 00
375021 3

LONGMAN Exam Dictionary. 2006. PEARSON.
Longman. ISBN-13: 978-1-4058-1860-5



Este obra está bajo una licencia de Creative Commons
Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional.

