

CHAPTER SEVEN

PROPERTIES AND INTERNAL
STRUCTURE OF THE LEXICON:
APPLYING THE GENERATIVE LEXICON
MODEL TO SPANISH*

ELENA DE MIGUEL

**7.1 Research on the Lexicon in the Early 21st Century
and its Impact on Second Language Teaching.
Brief Introduction**

Second language (L2) teaching is especially concerned with topics related to lexical acquisition, processing, and learning. It deals with aspects related to theoretical and experimental study, as well as classroom practice, which need new forms of explanation backed by theoretical advances.

Indeed, the L2 teacher faces problems related to words every day and often has no recourse other than arguing the authority of his native competence, if that is the case, and no solution beyond rote learning. A student with an acceptable level of L2 often cannot detect semantic polysemy of the kind found in the sentence (1a), whose subject could

* The research for this work has been partially funded by the research project *Electronic Multilingual Dictionary of Broad Sense Motion Verbs (Diccionario electrónico multilingüe de verbos amplios de movimiento)* (ref.: FF12009-12191; subprogram FILO), financed by the Ministry of Science and Innovation of the Spanish Government. I want to express my gratitude to Montserrat Sanz and José Manuel Igoa, editors of this volume, for their invitation to participate in it, and for their patience and support throughout the process of carrying out this work. I also want to thank Olga Batiukova, Gabriela Draghici, and Jacinto González Cobas for their interesting observations and suggestions regarding the work, and especially to Olga Batiukova for reviewing the final English version.

be the agent or the recipient¹ of the event of giving or taking class; nor understand a lexical ambiguity, as in (1b), where the word *historias* “stories” could allude to a story positively or negatively marked for the feature [\pm fiction] (thus causing different interpretations when combined with *contar*, “tell”). He might also have trouble to combine adequately the words depending on their sub-lexical features,² as in the inadequate examples (1c-d).³

In fact, even native speakers can have difficulty verbalizing the reasons why one combination of words is adequate and another similar combination is not; native speakers also commonly form apparently inadequate combinations, as in (1e-f):⁴

- (1) a. Hoy no voy a **dar clase** con Pedro
I will not have class with Pedro today
- b. No me gustan las **historias** que nos **cuenta** Pedro
I do not like the stories that Pedro tells us
- c. *Me he comprado unas **gafas morenas**
I have bought brown (=dark-skinned) sunglasses
- d. *Mis **profesores** de español son muy **bonitos**
My (male) Spanish teachers are very pretty
- e. *El Atlético es un reto y los **retos** son lindos de **correr**
*El Atlético is a challenge, and challenges are fun to run
(*EL PAÍS*, Summer 2003, “Deportes”, page 40)
*[Detuvieron a Michael Phelps] en **avanzado estado de embriaguez**
Michael Phelps was detained in an advanced state of drunkenness
(*EL PAÍS*, 10-11-2004)⁵

¹ Unless it is the same in their native language, which appears to reinforce the hypothesis of the determining importance of the properties of the L1 in the process of L2 acquisition. However, this idea is still debated, and there are different perspectives on it. Cf. Pérez Saldanya and Comajoan (2011) and references cited therein.

² This concept will be described later on, in section 7.2.

³ Examples taken from Sanz (2009) and (2011), which I will return to in the text, in section 7.5.

⁴ Note that some of these examples are acceptable in English but not in Spanish.

⁵ I have used an asterisk (*) for all examples whose lexical inadequacy allows judging them as “not belonging to standard language”. In reality, lexical violations seem to be different from syntactic violations, which is why in the examples (1c-f) other symbols could have been used: a hash mark (#), to show that they could be possible literary infringements, especially for poetic purposes; or a circle with an

In (1e), the inadequacy arises because *correr*, “to run”, is the predicate of *reto* “challenge”: challenges *se afrontan* “are faced” and *los riesgos* “risks” are run or taken. In (1f), *estado de embriaguez* “state of drunkenness” is modified by the adjective *avanzado* “advanced”, which is usually reserved in Spanish for other *estados* (“states”, either of gestation or putrefaction).

This work is based on the premise that inadequate combinations exist because the words have minor elements of meaning (in the form of features) that allow them to be combined with other words whose features are compatible. This premise implies the existence of lexical feature agreement processes that license the adequate combinations (*dar clase* “have class” with *alumno* “student” and with *professor* “teacher”), explain the different meanings of the same combination (*historias* “stories” as [± fiction] with *contar* “to tell”), and that exclude the word combinations whose features do not agree, as in (1c-f).

The analysis proposed to account for the generation and interpretation of the data as in (1), is part of a tendency followed in current linguistic studies to incorporate lexical information into explaining grammatical processes. In the words of Bosque (2004: CXXXI, my translation), “the analysis of the lexicon is now an inalienable part of the work of grammarians from a great number of formally, functionally, cognitively, or otherwise oriented trends and schools. All of these approaches recognize the extremely close relationship that exists between the content that a grammarian is supposed to explain and the content that a lexicographer is supposed to account for”. Indeed, if late nineteenth and early twentieth century linguists were recognized primarily for their work as phoneticians, and later on as phonologists (Anderson 1990), many late twentieth and early twenty-first century linguists and grammarians would believe that their work must incorporate lexical study and should attempt to identify the true reasons behind grammatical explanations within the lexicon.⁶

“x” in it, used by the Real Academia Española (Royal Spanish Academy of Language), to signal that the expression represents a violation of the lexical norms. I preferred to use the asterisk because, in the concept of language used here, both lexicon and syntax constitute components of grammar and both are computational and generative in nature; the infractions of the rules, operations, and principles of the lexicon always yield ungrammatical expressions in syntax.

⁶ For example, regarding the importance of the study of the lexicon for the field of language teaching, see Bartra (2009), Battaner (2009) and the works in García Platero and Castillo Carballo (2009), who approach the issue from very different perspectives.

The renewal of the interest in the meaning of words, relatively abandoned during a certain period of the discipline due to its idiosyncratic, heterogeneous, and boundless nature, traveled a long and complicated path,⁷ which has led to different proposals. These are generally characterized by decomposing the lexical units into low-level components with semantic and syntactic repercussions and searching for smaller and smaller entities with which to operate (De Miguel 2006).

However, there are many discrepancies in issues such as the distinction between what can be classified as world knowledge and what corresponds only to linguistic knowledge (that is, what does a speaker know because of his knowledge of the word, and what does he know because of his knowledge of the designated object or the event that it denotes?).⁸ There are also discrepancies regarding the nature of the relationship between lexical and syntactic information—an important question for much of the work being carried out regarding the *lexicon-syntax interface*, where the lexicon and the syntax come into contact. These are understood as two independent levels, with their own units, principles, relationships, and organization.⁹ Though proposals concerning the interaction between lexical and syntactic information vary among different models and trends, two different views responding to two different conceptions of the direction of this relationship can be recognized. One of these conceptions,

⁷ Referred to at times as “the lexicalist shift” (alluding to the Copernican shift).

⁸ For a brief argument in favor of the explanations that take linguistic knowledge into account exclusively or fundamentally see Moreno Cabrera (2003). Luque (2004) also deals with the difference between *realia* (world or object information) and *qualia* (information encoded in language, in words), together with an exhaustive list of phenomena that would interest someone studying the lexicon (from a typological perspective that takes into account processing and acquisition). Also see Lipka (1990), who is especially focused on the distinction between linguistic and extralinguistic (or encyclopedic) knowledge; the exhaustive introduction by Singleton (2000), that deals with complex lexical aspects of language conceived of not as a list of lexical units, but rather as a system of relations that determine a large part of what is habitually attributed to grammar; and the collection of works compiled by Peeters (2000), that deals with the crucial problem of the existence or not of a border between world knowledge and linguistic knowledge, as approached during the last decade in the studies of the lexicon. These are critically reviewed in the illuminating introductory chapter by Peeters (2000).

⁹ The statement that “all human language is structured into two basic components: a lexical component and a syntactic component” can be postulated as a semiotic universal (Moreno Cabrera 1997: 253, my translation). For the concept of “lexicon-syntax interface”, Mendikoetxea’s (2006) work is very comprehensive.

referred to as *projectionist*, posits that lexical information is projected to the syntax and therefore presupposes a bottom-up approach. In non-spatial terms, this view assumes that the properties and restrictions of lexical units determine the syntax. From this perspective, lexical semantics must account for the paradigmatic or structural relationships between words.

The other possibility, known by the term *constructionist*, posits that words are correctly interpreted only when they enter into certain syntactic contexts; therefore, the relationship is top-down. In non-spatial terms, this view defends the interpretation of words and their combinations as deriving from the structural properties of the syntactic configuration into which they enter. From this perspective, syntagmatic relationships restrict or determine meaning. This view is now predominant among generative linguists.¹⁰

This chapter opts for a projectionist view, represented by the *Generative Lexicon Theory* (hereafter, GL). In this model, the properties of words (their minimal content that takes the form of features) are responsible for licensing syntactic constructs, inasmuch as these materialize the meaning potential of the words. From this perspective, the context does not determine the meaning in the sense of generating it (which could lead to overgeneration); it instead visualizes non-transparent, sub-lexical features of the words. A construction may receive one interpretation or another, depending on the features that materialize; certain word combinations are unacceptable because their sub-lexical features do not agree. Therefore, the model adopted here may be defined as projectionist and compositional.¹¹ For our purposes, this model offers the necessary tools to explain some phenomena and data (such as those in (1)) that have not been accounted for in other analyses and that usually cause problems for L2 teachers and language teaching and learning specialists.

¹⁰ According to the definition proposed by Moreno Cabrera (2003: 18), the current state of research in theoretical grammar is characterized by a syntactically-based *grammatomorphism*. For a summary of the most used models, cf. Mendikoetxea (2009) and Val Álvaro (2010).

¹¹ This hybrid characterization is not free of controversy. It caused some authors to consider this model as projectionist, and others as constructionist: this is the case of Mendikoetxea (2009) vs. Mairal and Cortés (2009), in the same volume. See also De Miguel (2009).

7.2 The Generative Lexicon Theory: A Generative and Compositional Model of the Lexicon

The Generative Lexicon Theory (GL) is a formal theory of the organization and structure of the lexicon formulated by James Pustejovsky during the first half of the 1990s, and revised, expanded, and refined through many works (his own and those of his collaborators or followers, like Pustejovsky 1991, 1993, 1995; Pustejovsky and Boguraev 1993, 1994, 1996; Pustejovsky and Bouillon 1996; Pustejovsky and Busa 1995; Pustejovsky and Rumshisky 2008, 2009), during the years since its initial publication.¹²

Within this model, the lexicon not only constitutes a simple repository of static definitions but is also given generative power: it is a level organized according to a rich and recursive theory of meaning decomposition, which encompasses a large part of the meaningful and creative potential of the language. The model assumes a series of general organizing principles for the lexicon and a set of mechanisms that affect the definitions of words and that may account for their contextual polysemy.

GL is a lexical-semantic theory of a generative and compositional nature. It is a generative model of the lexicon, because it aims at accounting for what constituted the foundational question of the Chomskyan program: How is it possible to generate a non-finite number of results through a finite number of resources? In our case, the question is how a potentially infinite range of meanings is obtained, depending on the context, from a limited number of senses listed in the lexicon and a finite number of principles, operations, and mechanisms; not only how the speaker can generate new senses, but also how the listener can decode them. Word combinations are crucial for explaining which mechanisms allow a lexical form to acquire a unique and specific interpretation within a sentence. Hence, syntactic context plays an important role in generation and decoding of meaning. The GL model is thus compositional.

The basic goal of GL is to address the problem of why a word, not only in different contexts but also within the same context, carries different meanings, as in the classic example in (2a), which could be interpreted as “I began reading the novel” or “I began writing it”. This

¹² Many of them are included in the bibliography, from works where the premises of the model were advanced (in 1988) up to some of the most recent works. The classic version of the model is presented in Pustejovsky (1995). De Miguel (2009) reviews the latest developments of the basic premises of the theory.

constitutes a typical case of what is known as *logical* or *systematic polysemy*.

- (2) a. He empezado la novela
 (“He empezado a {leer/escribir} la novela”)
 I began the novel
 I began {reading/writing} the novel
- b. La clase no favorece la discusión
 (“{La lección impartida/el alumnado/el aula} no favorece la discusión”)
 The class does not favor discussion
 {The lesson given / the group of students /the classroom} does not favor discussion
- c. Hablé a los alumnos de lingüística
 (“Hablé [a los alumnos] [de lingüística]” / “hablé [a los alumnos de lingüística]”)
 I talked to the students of linguistics
 I talked [to the students] [about linguistics] / I talked [to the linguistics students]
- d. El cuadro de Inés me encantó
 (“El cuadro {que Inés pintó/en que Inés sale/que Inés posee} me encantó”)
 I loved Inés’s painting
 I loved the painting {that Inés painted / where Inés is portrayed / that Inés owns}

In principle, the ambiguity of (2a) does not seem to be a consequence of lexical polysemy, *i.e.*, ambiguity of the words *novela* “novel” or *empezar* “to begin”. That does seem to occur in (2b), whose different meanings may be attributed to the different senses of *clase*, “class”, listed in a standard dictionary: the noun of an event (“the lesson given”), a collective noun (“the group of students”) and the noun of an object (“physical space synonymous with *aula* [classroom]”), whose conditions of sonority, lighting or temperature could impede debates or delving deeper into a topic. The ambiguity of (2a) does not seem to derive from structure either, as it only receives one syntactic analysis, unlike the classic example in (2c), where two different analyses justify two possible interpretations. Finally, the polysemy of (2a) does not appear similar at first to that of (2d): *cuadro* “painting” is a predicative noun (it is iconic and also an artistic creation) that materializes a single argument (*de Inés*

“of Inés”) as a complement, which could be assigned the thematic role of agent, patient, or possessor.

I return below to (2d) and (2b) to defend that their multiple interpretations come from the same source as those in (2a), but for now, we assume that this last example, when seen from a traditional perspective, cannot be attributed to any of the three causes generally known as *lexical*, *structural*, and *semantic ambiguity* or *polysemy*.

To approach the problem of logical polysemy, as in (2a), GL adopts two premises that determine the character of the model and that make it similar to other recent models of lexicon. These two premises are the *decomposition of meaning of the lexical units* and the *compositional nature of lexical meaning*. The first premise assumes that words have an internal or sub-lexical structure. That is, they do not constitute atomic entities but rather meta-entries whose meaning may be broken down into different sub-lexical features. As far as the semantic compositionality is concerned, the idea is that there exists a set of regular agreement principles and mechanisms which operate on the sub-lexical features.¹³ Based on a minimal word definition, these can generate multiple new meanings for a word when it enters into syntactic composition with other lexical units.¹⁴ Decomposition and compositionality are therefore two non-contradictory premises.

As mentioned above, the main goal of GL is to demonstrate that words can acquire multiple meanings depending on the context in which they appear, which is a general and frequent phenomenon in languages, as illustrated in (3):

- (3) a. Una maleta ligera; una comida ligera; una comedia
 ligera
 A light suitcase; a light meal; a light comedy

¹³ The term *agreement of sub-lexical features* is borrowed from Bosque (2004). Bosque (2004) is actually, as its director defines it, a dictionary of lexical concordances or redundancies.

¹⁴ A traditional distinction between lexical and functional categories defined the former as bricks of a building joined by the cement made up by the latter (Bosque 1989). The conception of lexical words as non-atomic entities works better with Cohen’s metaphor (1986) that describes them as sacks whose meanings vary depending on the others with which they were joined, whose size and shape are molded in the combination. In De Miguel (2009), I added to this metaphor that, for Pustejovsky’s hypothesis, the sacks must be porous, because through these pores, particles of meaning enter and leave from the contiguous sacks, causing “lexical reactions” that alter the initial definitions of the words.

- b. Cortar el césped; cortar el pastel; cortar el gas
Cut the grass; cut the cake; cut off the gas
- c. Una brisa ligera; un combustible ligero
A light breeze; a light fuel
- d. Cortar una relación; cortar un dedo
End a relationship; cut a finger
- e. Luis se ha cortado el dedo con el cuchillo del jamón y tiene una herida bastante profunda / Luis se ha cortado un dedo con la motosierra y lo ha metido en hielo a la espera de la ambulancia
Luis cut his finger with the ham knife, he has a fairly deep wound / Luis cut his finger off with a jigsaw and put it on ice while waiting for the ambulance
- f. #Una laguna ligera; #cortar el sol¹⁵
#A light lagoon; #cut the sun

In (3a), the adjective *ligera* “light” has different meanings depending on the noun that accompanies it. Thus, *ligera*, when predicated of *una maleta* “a suitcase”, means that it is not bulky; as the predicate of *una comida* “a meal”, it means that it is easy to digest; and with *una comedia* “a comedy”, it means that the comedy does not require reflection on the part of the viewer. Similarly, *cortar* “to cut” in (3b) denotes different events when it is the predicate of *el césped* “the lawn” (in which case it is similar to *recortar* “to cut down”), *el pastel* “the cake” (where it could be changed for *trocear* “to cut up”) or *el gas* “the gas” (where it would mean *interrumpir*, “to interrupt service”).

Interestingly enough, these examples do not exhaust all the possible cases: for example, the meaning of *ligero*, when predicated of *brisa* “breeze” or *combustible* “gas” in (3c), is different from (3a). The same may be said for *cortar*, which is not interpreted as “recortar,” “trocear,” or “interrumpir” when it has other complements, like *una relación* “a relationship” or *un dedo* “a finger” in (3d). In fact, *cortarse un dedo* (to cut one’s finger) is in itself polysemous, as it can mean “hacerse un corte en un dedo” (to make a cut in one’s finger) or “seccionárselo por completo” (to cut one’s finger off), as in (3e).

This typical property of words, which allows them to yield multiple, probably infinite meanings when combined, forces the lexicographer to create extensive dictionary entries. These entries are necessarily incomplete, and consulting them can be an extremely complex and

¹⁵ Here, the dash symbol was used instead of asterisk to signal that these examples are acceptable with a metaphorical interpretation in an “exceptional” context: for example, if understood as an infraction typical of poetic language.

somewhat useless task for a non-native speaker. In L2 teaching practices, including writing manuals and creating classroom practice activities, searching for general principles or rules is often abandoned, and a long, arbitrary, incomplete and idiosyncratic list of cases is instructed to be memorized. For the lexicographer and the semanticist, data in (3) have sometimes been the motive behind an extreme position: denying the existence of meaning independent from use.

Light verbs, including *dar* “to give” or *hacer* “to do” or “to make”, illustrate this complicated behavior: their multiple meanings require that standard dictionaries include very extensive entries that do not list all possible definitions, nor is a “real” meaning of these verbs established “out of context”. Consulting the definitions is thus complex and not practical. In the practice of L2 teaching, related languages behave differently when selecting for these verbs, which contributes to the deepened awareness of idiosyncrasy and chaos. These sensations have burdened the research on lexicon, often regarded as an irregular and random component that escapes systematic study. For an L2 teacher, it is difficult to explain (and certainly discouraging, as far as the search for common principles and mechanisms is concerned) how come in Spanish, walks are “given” (*los paseos se dan*), as are kisses (*los besos se dan*), and caresses are “done” (*las caricias se hacen*); by contrast, in Italian, a closely related language, walks are “done”, unlike kisses, which are “given”, while caresses may be “done” or “given”.

Light verbs are frequently studied in the recent bibliography because they play an important role in determining classes of words.¹⁶ Contrary to the traditional distinction, which attributes the predicating function to the verb, light verbs (whose meanings vary depending on the event nominal that accompanies them, and with which they form a predicative unit and share the arguments) delegate part of this function to the noun. Conversely, with a referential noun, the same verb functions as fully predicative, as it maintains the basic or literal meaning; this is illustrated by the contrast in (4). The examples of (4a) are ungrammatical because a light verb must share the subject argument with the event nominal. The examples in (4b) are acceptable because the object is not interpreted as being created through the verbal event (as in *Juan dio una respuesta* ≈

¹⁶ In fact, the light verbs class not only has been increased, as in Bosque (2001), who included the so called *heavier light verbs*, but also its very nature has been questioned, as in De Miguel (2011), where it is argued that any verb can be a light verb depending on its combination with a more or less predicative complement, like with *abrir {la puerta/la sesión}* (“open the door/the session”), an analysis advanced in De Miguel (2010).

“respondió” [Juan gave an answer = “he answered”]) but is instead referential: it is a pre-existing object reproduced by a subject (the verb’s subject), which differs from the agent who created it.

- (4) a. Juan dio {un/*mi} paseo por el parque / Pedro hizo {una/*mi} caricia
 Lit.: Juan gave {a/my} walk through the park / Pedro did {a/my} caress
 Juan took {a/*my} walk through the park / Pedro gave {a/*my} caress
- b. Juan dio mi respuesta / Pedro hizo mi análisis
 Juan gave my answer / Pedro did my analysis

This work is interested in how light verbs prototypically illustrate a more general phenomenon, which we have seen in (3): the words’ meaning changes depending on the combinations into which they enter. On the one hand, the data support the compositionality premise; on the other hand, the restrictions for combining light verbs with event nominals (**hacer una respuesta/*dar una pregunta* “*to make an answer/*to give a question”) imply that word combinations are not completely free. This supports the decomposition premise. If we assume that words have an internal sub-lexical structure, we begin to realize why certain combinations are possible while others are not. By postulating general processes for the agreement of sub-lexical features (which license or reject combinations and even favor extensions of meaning in supposedly non-legitimate combinations), we offer an explanation that generalizes or normalizes a supposedly idiosyncratic or random behavior. In pedagogical and lexicographical practices, this hypothesis offers interesting applications.

Many verbs behave differently, depending on whether their complement pre-exists or not, not just the light verbs (as in (4a) vs. (4b)). In (5), other examples show how a verb denotes a creation event when combined with a non-pre-existing object (*abrir un pozo* “open a well”; *lanzar una acusación* “make an accusation”; *levantar acta* “take the minutes”), and how the same verb implies a change of state event when combined with a pre-existing object (*abrir una puerta* “open a door”; *lanzar una jabalina* “throw a javelin”; *levantar la maleta del suelo* “pick the suitcase from the floor”; *romper un jarrón* “break a jar”). If the change of state denoted by the verb is semantically incompatible with the sub-lexical features of the pre-existing object, another change of state interpretation emerges: that of no longer existing (*levantar un embargo* “lift an embargo”; *levantar el novio a una amiga* “steal a friend’s boyfriend”; *romper una relación* “end

a relationship”). The three interpretations respond to a regular pattern, which can also be applied to more metaphorical or idiomatic cases, and to *cortar* in (3d) and (3b)¹⁷:

- (5) a. abrir {una puerta / una lata de conserva / un pozo / una sesión}
to open {a door/ a jar of preserves / a well / a session}
- b. lanzar {una jabalina / una acusación / un libro / una promoción / un beso}
to throw a javelin/ to hurl an accusation/ to launch a book/ to launch a promotion/ to launch a kiss
- c. levantar {una maleta del suelo / acta / sospechas / un embargo / el novio a una amiga}
to lift a suitcase from the floor / to take the minutes / to arouse suspicions / to raise an embargo / to steal a boyfriend from a friend
- d. romper {un jarrón / una relación}
to break a jar / to end a relationship

The surprising regularity that underlies the apparent interpretative chaos of the data in (5) and (3) is a piece of good news for the lexicographer, the teacher, and the learner of an L2 lexicon. It is also extremely interesting for a linguist from the theoretical perspective, as the licensing and interpretation of the combinations illustrated in (1) through (5) highlight lexical properties that allow these words to be adequately combined amongst themselves, and augment, reduce, or modify their meaning in supposedly non-legitimate combinations. To demonstrate this, the lexicologist must make use of grammatical models that include dynamic, flexible, and context-sensitive lexicons. Assuming this premise about the nature of the lexicon simplifies the L2 teaching, too.

As seen above, combining words is not absolutely free, which indicates that lexical meaning (even if it is minimal) exists independently from the use of the words in context. This is why GL postulates that,

¹⁷ Of course, not all combinations of event nominals and change of state verbs yield the interpretation of “pasar a existir” (come into existence, e.g., *levantar acta* (“taking minutes at a meeting”) or *dejar de existir* (“cease to exist”, e.g., *levantar un embargo* (“lift an embargo”). For instance, **levantar un problema* (“raise a problem”, in the sense of “start existing”) or **levantar un conflicto* (“cease a conflict”, in the sense of “stop existing”) are not available. This means that the sub-lexical features of words in those combinations are not compatible and do not fulfil the basic requirement of lexical agreement, which is responsible for triggering different interpretations.

though it may be constructed within the context, meaning is not generated freely and without restrictions, it is always lexically motivated: the “new” meanings are contained as potentialities in the word definition in the lexicon. This explains the unacceptability of the expressions in (3f), as opposed to (3a-e), which can be generated and interpreted successfully.¹⁸

7.3 Basic Assumptions of GL: The Information Contained in Lexical Entries

7.3.1 Underspecification

The dynamic, generative, and compositional conception of the lexicon described above is based on the assumption that word definitions in the mental lexicon are underspecified. They can therefore acquire more precise or specific meanings in different contexts, as pointed out in (6):¹⁹

- (6) Underspecification:** A lack of specification in the lexical entries that allows them to intervene in different syntactic structures and, as a consequence, in different operations of semantic composition.

If, following (6), the lexical level contains underspecified lexical entries that can subsume multiple senses of a word within a context, listing them becomes unnecessary. The relationships that they maintain become evident, resulting in a smaller and more predictable mental lexicon.

Words that have underspecified definitions are specified or determined within a context, when they are combined with others whose sub-lexical features agree. This is achieved through mechanisms that allow materializing their semantic potentialities. Such is the case of *ligera* “light” as the predicate of *maleta* “suitcase”, as *maleta* is a noun that refers to an object [+container] with weight and volume. As *maleta* has the sub-lexical feature [container], it can enter into the syntactic structure in (7a), unlike *comedia* “comedy” in (7b) and *laguna* “lagoon” in (7c):

- (7) a. La maleta _[+ container] está llena
The suitcase is full

¹⁸ Therefore, it is claimed here that GL is a projectionist model.

¹⁹ According to Geeraerts (2002: 28), GL inherits the generativist ideal of formal semantic representation put forth in Katz (1972) but, instead of basing itself on a static formalism, it introduces a logical formalism in a flexible model, which is why the author defines it as “neo-generativist”.

- b. *La comedia _[-container] está llena
 *The comedy is full
- c. *La laguna _[-container] está llena
 *The lagoon is full

When the features of the combined words do not agree, two possibilities arise. The first one is that the resulting expression is not legitimate; it provokes an interpretive collapse. Such is the case of *una laguna ligera* “a light lagoon” in (3f). Secondly, a rescue mechanism may intervene to recategorize the features of one of the disagreeing words, making its meaning compatible with the other word so that both constitute an interpretable expression. Such is the case of *una comedia ligera* “a light comedy” in (3a): as *comedia* is negatively specified for the sublexical feature [\pm container], *ligera* cannot modify this feature. It can, however, modify another feature that refers to the [density] of the content. As *laguna* is negatively specified for the features [\pm container] and [\pm density], there is no possible rescue. The sub-lexical features that agree (automatically or through a process of matching) and allow for or exclude combinations of (3) are contained within the underspecified word definitions, which license the possible combinations and the interpretation of the resulting expression.

Acknowledging that these processes exist implies also assuming that the word meaning does not constitute an atomic and closed definition. Rather, it contains different structured and overlapping kinds of information which interact in different syntactic combinations (this is what the model argued for here defends). These bits of sub-lexical information are structured in different levels of representation which make up meta-entries. The meta-entries, combined with the generative mechanisms that operate on them, allow reducing the size of the mental lexicon.

7.3.2. Levels of Representation

GL assumes the existence of four levels of representation, which structure the information contained in the lexical entries: the *Argument Structure*, the *Event Structure*, the *Qualia Structure*, and the *Lexical Typing Structure*.

7.3.2.1. *The Argument Structure (AS)* contains the specification of the number of the arguments of a predicate, the semantic class to which they belong (for example, [individual], [object], [event]), and how they are

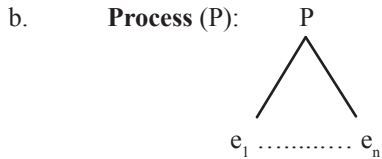
realized syntactically. This is a well-known concept in syntax, although Pustejovsky introduces some interesting new distinctions.

7.3.2.2. *The Event Structure (ES)* defines the event type of the predicate (a word or a phrase). According to Pustejovsky, there are three types: *state*, *process*, and *transition*. Their internal or sub-event structures are represented in the tree diagrams in (8).

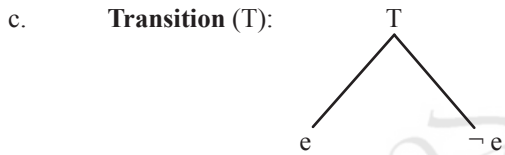
(8) *Event types according to Pustejovsky (1991, 1995)*



A single event, which is evaluated relative to no other event: *amar* (to love), *saber* (to know).



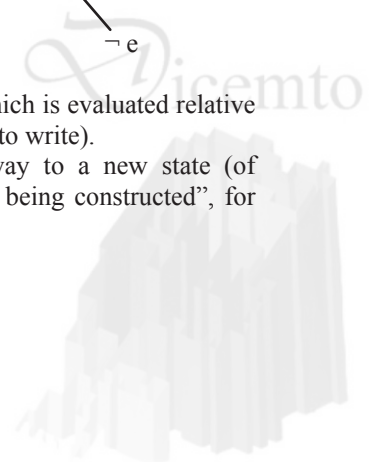
A sequence of events identifying the same semantic expression: *correr* (to run), *nadar* (to swim).

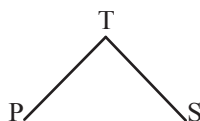


An event identifying a semantic expression, which is evaluated relative to its opposition: *construir* (to construct), *escribir* (to write).

As transition implies a process that gives way to a new state (of “something not being constructed” to “something being constructed”, for example), it may also be represented as follows:

²⁰ The letter *e* is the variable for any type of event.



d. **Transition (T):**

The main appeal of this proposal is that it assumes that events are not atomic entities, but that they have internal structure and may therefore be decomposed into phases or subevents. This explains how certain morphological and syntactic processes focus on only one sub-lexical phase, as in (9): (9a) accepts the durative and punctual adverbials because the predicate contains two phases; (9b) only accepts the punctual adverbial because the predicate lacks the process phase (P); (10) is ambiguous because the adverb may modify the moment at which the event starts (S) or the phase in which the event is developed (P), independently from how it starts:²¹

- (9) a. Isabel construyó su casa {en 1998 / durante dos años}
 Isabel built her house {in 1998/ for two years}
 b. Isabel llegó {a las diez / *durante dos años}
 Isabel arrived {at ten/*for two years}
- (10) a. El péndulo osciló bruscamente
 “Bruscamente el péndulo se puso a oscilar” / “el péndulo describió una oscilación brusca mientras osciló”
 “Abruptly the pendulum started to oscillate” / “the pendulum described an abrupt oscillation while it oscillated”

7.3.2.3. *The Qualia Structure (QS)* is perhaps the most novel and interesting proposal of GL as far as levels of representation are concerned; it is also the one that yielded most productive applications.²² It assumes that the underspecified definition of the words in the lexicon includes information about fundamental characteristics of the entity to which it

²¹ This analysis implies that the ES of *oscilar* is [S + P], a structure not postulated in Pustejovsky’s model, but predicted in a geometric model of events (see De Miguel and Fernández Lagunilla 2000, 2006, among others, for an extension of the ES proposed in the GL model).

²² As a matter of fact, the most important theoretical consequences, and lexicographical and computational applications of Pustejovsky’s model refer to QS, whose assumptions appear even in authors that do not subscribe to other principles and mechanisms of GL, such as Cruse (2004), Jackendoff (2002) and Van Valin (2005), according to Batiukova (2008).

refers (an object or an event), of the type “how it came to exist”, “what is its internal constitution”, “what is it used for” or “how is it formally different from other objects in a more extensive domain”. This information is structurally encoded and hierarchized according to four elements of meaning called roles or *qualia*: agentive, constitutive, telic, and formal.²³

The *qualia* definition is given in (11), and (12) exemplifies how adjectival and prepositional complements of a noun can materialize or make linguistically explicit different sub-lexical information of its QS:

- (11) a. **Formal *quale***: distinguishes the object within a larger domain (that is, information about the orientation, magnitude, shape, color, dimensionality, and position).
 b. **Constitutive *quale***: encodes the relationship between an object and its constituents, or proper parts (that is, information about the material, weight, parts and component elements).
 c. **Agentive *quale***: encodes factors involved in the origin or “bringing about” of an object (that is, information about the creator, the artifact, the natural kind or the causal chain).
 d. **Telic *quale***: encodes the purpose and function of the object (that is, information regarding the purpose that an agent has in performing an act or a built-in function or aim which specifies certain activities).
- (12) a. pista {rojiza, cubierta, rectangular} [formal *quale*]
 {reddish, indoor, rectangular} court
 b. pista {de hierba, de cemento, de hielo}
 [constitutive *quale*]
 {grass, cement, ice} court
 c. pista {artificial, municipal, de diseño} [agentive *quale*]
 {artificial, municipal, designed} court
 d. pista {de baile, de tenis, de patinaje} [telic *quale*]
 {dance, tennis, skating} court

²³ The four *qualia* of Pustejovsky are directly inspired by the four *aitiai* or modes of explanation that Aristotle proposes in his *Physics*, as Moravcsik (1975, 1991) pointed out. The difference is that the latter are ontological distinctions within a proposal of explanation of the interpretation of real-world objects, and those of Pustejovsky are linguistic distinctions encoding the properties of objects in the real world. This distinction should be taken into account, since critiques regarding the incoherence of the model are frequently made because of the amount of encyclopedic information included in QS.

7.3.2.4. *The Lexical Typing Structure* establishes how one word relates to another in the mental lexicon using the information contained in the QS.²⁴ Let us see a simple example taken from Pustejovsky (1995):

- (13) a. Novel: [book] [narrative]
 [created through writing]
 [destined for reading]
 b. Dictionary: [book] [listing of words]
 [created through compiling]
 [destined for consultation]

The nouns *novel* and *dictionary* allude to the same type of object; that is, they share the information coded in the formal *quale*. Specifically, they are both formally defined as [book]. However, their purpose (information contained in the telic *quale*) is different: *novel* is an object [book] destined [for reading], while *dictionary* is an object [book] destined [for consultation].

This difference has consequences for the interpretation of syntactic structures. For instance, *I began the novel* in (14a) may imply “I began writing it” or “I began reading it”, while *I began the dictionary* in (14b) only receives the first interpretation:

- (14) a. I began the novel [“I began {writing / reading} it”]
 b. I began the dictionary [“I began compiling it”]

The examples from (14) show how different levels of representation are involved in determining the global meaning of the linguistic expressions. The noun *novel* is defined in its QS as an object that is destined to be read and that is created through writing, which makes (14a) ambiguous. However, (14b) is not ambiguous (the interpretation “I began consulting the dictionary” is not available). The reason stems from the ES of the verbs *to consult* and *to read*. While *to read a novel* is an accomplishment (a durative event whose ES is represented in (8d), <T [P+S]>), *consulting a dictionary* is a punctual event that lacks the P phase. It is therefore incompatible with the meaning of the verb *to begin*, just like

²⁴ According to Val Álvaro (2010: 26, fn.6), the information at this level could be considered, in general, as part of the formal *quale* of the QS.

any other punctual verb: **to begin arriving*.²⁵ Examine, lastly, the examples of (15):

- (15) a. He empezado el periódico
 (“he empezado a leer el periódico”)
 I began the newspaper (“I began reading the newspaper”)
- b. Hemos empezado el periódico (“hemos empezado a {leer/redactar} el periódico”)
 We began the newspaper (“we began {reading / editing} the newspaper”)
- c. Me he empezado la novela (“he empezado a leérmela”)
 I began the novel (“I began reading it”)

In (15a), there is only one interpretation, though *newspaper* has information relating to two events (the agentive and the telic). This happens because *newspaper* necessarily implies a process of collective creation (which is not the case for *novel* or *dictionary*); thus, the two interpretations are only recovered in the plural (15b). Finally, (15c) only has one interpretation (which is not the case for (14a)) because the clitic *me* is a culminating aspectual operator that is only compatible with events that have culminated and given way to a new state, requiring a pre-existing object (its meaning is similar to certain resultative phrasal verbs in English: *beberme algo* = “drink something up”). This information is only implied in the event encoded in the telic *quale*; the agentive *quale* does not assume the pre-existence of the object.²⁶

7.3.3. Types of Words by their Sub-Lexical Structure

Pustejovsky distinguishes the following types of nouns depending on the way in which QS information is combined in their meaning.

7.3.3.1. *The natural types*: These words are not polysemous. Their meaning is derived from the information contained in the formal and constitutive roles: this is the case of nouns such as *caballo* (“horse”), *laguna* (“lagoon”), or *sol* (“sun”), which are entities belonging to a certain

²⁵ Of course *empezar a llegar* “to begin arriving” and *empezar a consultar el diccionario* “to begin consulting a dictionary” are legitimate expressions if referring to a habitual event, that is, the repetition of a punctual event.

²⁶ For the analysis of this aspectual clitic in Spanish, see De Miguel and Fernández Lagunilla (2000).

category and which are defined through their constitution (external and internal).

7.3.3.2. *The unified or functional types*: These words denote created entities or artifacts. They contain the information encoded by natural types and in addition include the elements of meaning referred to by the agentive and telic roles, *i.e.*, how an object was created and what its function is. This is the case of *biberón* (“feeding bottle”), *cuchillo* (“knife”), or *profesor* (“teacher”); these are all physical objects and instruments that have certain functions (drinking, cutting, teaching). In this sense, the unified types have more dimensions than simple types, as they incorporate more defining information. In theory, they are not polysemous, but in certain contexts they may be ambiguous because they combine values from different QS roles. For example, *biberón* is defined in the QS as a “physical object, not natural, usually made of plastic, that can contain liquid and is used for drinking”. When put into different contexts, one or more elements of this definition may materialize, as in (16a-b); in a sentence like (16c), it may also lead to polysemy, because the verb *dar* (“to give”) may be combined with information encoded in either the formal or the telic role:

- (16) a. El biberón se rompió [= “el objeto de plástico se rompió”]
 The feeding bottle broke [= “the plastic object broke”]
- b. El niño se tomó el biberón [= “el niño se tomó la cantidad de líquido que contenía el recipiente biberón”]²⁷
 The child drank the bottle [=“the child drank the amount of liquid that the feeding bottle contained”]
- c. La abuela dio el biberón al niño [= “la abuela dio al niño el objeto de plástico” / “la abuela dio de beber al niño la cantidad de líquido que contenía el recipiente biberón”]
 The grandmother gave the child the bottle [=“the grandmother gave the plastic object to the child” / “the grandmother fed the child the amount of liquid the feeding bottle contained”]

²⁷ As we will see right away, the metonymic interpretation of (16b) is obtained through regular and general mechanisms of generation of literal meanings. It should be pointed out that the appeal of the GL model is due in part to its potential for demonstrating how certain metaphorical or metonymic extensions are generated without the necessity for postulating exceptional mechanisms, or increasing to infinity the size of the mental lexicon.

The difference between natural and unified types is linguistic in nature, *i.e.*, it is related to the words, not to the real world entities the words refer to. As noted in De Miguel (2009), although a natural type, for example, *caballo* “horse”, from the cultural or encyclopedic perspective, may have a prototypical function (a horse as an entity in the world prototypically performs the function of a draught animal), it does not mean that this information forms an inherent part of the definition of the word *caballo*. The definition of the entity of a horse is independent from whether or not it performs this function. However, the function of an instrument is a fundamental part of its definition: a utensil that is no longer used and becomes a mere decorative object (for example, a glass feeding bottle) is still an object created for a certain purpose. This is why we can say *un antiguo biberón* “an old feeding bottle, a feeding bottle that is no longer used as such”; **un antiguo caballo*, “an old horse”, meaning “a horse that is no longer used with whatever purpose”, on the other hand, is not acceptable. In this sense, there are certain information or roles in the QS that are determining and obligatory, and there are others that are weak and dispensable.²⁸

The difference between the natural and unified types may be illustrated with the examples in (17). Though (17a) is perfectly interpretable (“the child finished drinking the bottle”), a very specific context is needed for (17b) to be interpreted (*e.g.*, manual work or a painting class: it requires recategorizing the word *horse* as a unified type):

- (17) a. El niño acabó el biberón
 The child finished the bottle
 b. El niño acabó el caballo
 #The child finished the horse

Pustejovsky emphasizes, however, that this type of recategorization and the conjunction of different values in unified types are not free or unrestricted processes. They are restricted by the general principles that license when and how one type of word “inherits” the properties from other types of words.

²⁸ See Pustejovsky (2006, 2008). The different weight of pieces of information in the QS of words has consequences for their lexicographical definition, and in fact, it is habitual practice during the creation of dictionaries to take into account some particular content depending on the type of entity described (*cf.* González Cobas 2010). For lexicographical applications of the GL model, see Batiukova (2008).

7.3.3.3. *The complex types* (also known as *dot objects*, objects formally represented by the symbol “•”) are objects composed of two or more types in their QS. They constitute a Cartesian product (x,y) of the types that compose them. To represent them formally, the logical operator *dot* (•) constructs complex types (“a•b”) from two simple types (“a” and “b”). For instance, [food] and [event] are joined for a *meal* ([food] • [event]), a complex type formed by two apparently incompatible or contradictory simple types.

- (18) a. *comida* “meal”: ([food] • [event])
 b. *libro* “book”: ([physical object] • [content])

Constructing a complex type results in a word that is “systematically” polysemous, such as *meal*, *lecture*, *construction* or *book*, which simultaneously contain at least two senses in their formal *qualia*. In certain contexts, different interpretations are projected simultaneously; in other contexts, one interpretation is selected disjunctively.²⁹ For example, *book* is specified in its formal *quale* as [physical object] and [information], so the sentence (19a) is systematically ambiguous. A similar analysis accounts for (19b) and (19c), which refer to contexts that evoke simultaneously different meanings of *lecture* and *construction*, as [object] and [event]:

- (19) a. No me gusta el libro [= “no me gusta {su formato / su contenido}”]
 I do not like the book [= “I don’t like {its format / its content}”]
 b. No encuentro la conferencia [= “no encuentro los papeles” / “no sé dónde tiene lugar el evento”]
 I can’t find the lecture [= “I cannot find the papers” / “I don’t know where the event is taking place”]
 c. La construcción no agrada a los vecinos [= “no agrada {el edificio / el que tenga lugar el evento de construir}”]
 The neighbors do not like the construction [= “they don’t like {the building / that the event of constructing is taking place}”]

²⁹ The complex types proposal offers an interesting account of ambiguous nominalizations with an event or a result reading—previously analyzed in terms of aspectual information (cf. Grimshaw 1990, for example)—without requiring additional stipulations other than the GL general principles and mechanisms.

In some contexts, only one possible interpretation is chosen. For example, if the complex nominal object is combined with a verb that requires a simple type as a complement (*libro* “book” with *quemar* “to burn” in (20a) or *construcción* “building” with *demoler* “demolish” in (20b)), only the object interpretation is activated. The verbs *durar* “to last” and *tener lugar* “take place”, conversely, only activate the event interpretation, as shown in (20c-d):

- (20) a. El inquisidor quemó el libro
The inquisitor burned the book
b. La construcción fue demolida
The building was demolished
c. La conferencia tuvo lugar a las diez
The conference took place at ten
d. La construcción duró mucho
The construction took a long time³⁰

7.4 Generative Mechanisms

Finally, let us examine the lexical agreement mechanisms that operate on the information contained in the underspecified representation of the words to determine whether some combinations are legitimate, reject others as uninterpretable, and intervene to rescue other supposedly non-legitimate combinations. In the latest versions of the model (Pustejovsky 2008), there are five mechanisms, which are presented below.

³⁰ In previous sections I illustrated these types of words using nouns, but the distinction also applies to verbs: *llover* “to rain” may be considered a simple type, *escribir* “to write” a unified type (which in certain contexts activates one or more of its phases: *I am writing a book* / *I have written a book*) and *oscilar* “oscillate”, as seen in (10a), a complex type ([state] • [process]), it can have different interpretations in the same context. There are, however, other possible interpretations: Pustejovsky (2006) argues that a complex event is one in which at least one complex type argument participates. For instance, *ver una película* “to see a movie” is a systematically ambiguous predicate, because *película* “movie” is a complex type ([story] • [physical object]), that shows different syntactic behavior depending on which formal *quale* materializes. As pointed out in De Miguel (2010), *ver una película* accepts passive, while *ver una película* does not (*la película fue vista por muchos espectadores* “the movie was seen by many spectators” / **la película fue vista encima de la mesa* “*the movie was seen on top of the table”). According to Murphy (2010), this is a particularly attractive aspect of GL, because it allows for a unified treatment of grammatical categories from the perspective of lexical semantics.

7.4.1. Selection

Selection (also called *pure selection*) does not require any special adjustment operation; it applies when a type required by the predicate is clearly satisfied by its argument. If the information contained in the combined words' QS is compatible, selection is activated to license the combination and its interpretation. The examples (21 a-b) illustrate this.

- (21) a. *Aventurar una conjetura*³¹
 To venture a guess
- b. *Pintar un cuadro*
 To paint a painting
- c. *La conjetura de Luis* (= “la conjetura que Luis aventuró”)
 Luis’s guess (= “the guess that Luis ventured”)
- d. *El cuadro de Renoir* (= “el cuadro que Renoir pintó”)
 Renoir’s painting (= “the painting that Renoir painted”)
- e. *El cuadro fue pintado* *({por Velázquez / en 1618 / al óleo})
 The painting was painted *({by Velázquez / in 1618 / in oils})
- f. *La conjetura fue aventurada* *({por Luis / a la ligera})
 The guess was ventured *({by Luis / without thinking})

The full compatibility between the features of the noun and the verb in these combinations has interesting consequences: if the verb and the noun share content to the point of constituting redundant expressions, the verb may be omitted. Because the meaning of *conjetura* “guess” implies the meaning of *aventurar* “to venture”, and that of *cuadro* “painting” implies that of *pintar* “to paint”, *la conjetura de Luis* “Luis’s guess” in (21c) can be interpreted as “the guess that Luis made or ventured” and *el cuadro de Renoir* “Renoir’s painting” in (21d) as “the painting that Renoir painted”.³² The redundancy between the noun and the verb also explains why the passive, as in (21e-f), should be avoided, unless a phrase appears that can make the passive construction predicative and informatively

³¹ Example taken from Bosque (2004).

³² Amongst other possibilities, of course: given that *painting* is a complex object that has two pieces of information in its formal *quale* ([iconic object]) and [physical object]), *el cuadro de Renoir* “Renoir’s painting” can also be interpreted as “the painting that Renoir is in” or “the painting that Renoir possesses”. This ambiguity, regularly explained in terms of thematic roles, was reformulated in terms of the QS in De Miguel (2009, 2010).

relevant. This is the case of *en 1618* “in 1618” or *a la ligera* “without thinking”.³³

7.4.2. Accommodation

Accommodation is a non-canonical (“not pure”) mechanism that operates when the argument features do not clearly agree with the predicate features but may do so in a subsidiary way, if the argument is a hyponym of the type selected by the predicate, as in (22).

- (22) a. oír la música
hear the music
b. oír _____ [+ sound]
hear _____ [+ sound]
c. música < sound >
music < sound >

In this case, because of lexical type inheritance, the argument gives the predicate access to the type it requires. *Oír* “to hear”—“to perceive sounds with the ears”, according to *Diccionario de la lengua española de la RAE* (*Spanish Language Dictionary of the Royal Spanish Academy* (DRAE, 2001))—is a verb that requires a complement with the sub-lexical feature [+sound]; *un grito* “a yell”, *un ladrido* “a bark”, are just that, therefore *oír un grito* “to hear a yell” or *oír un ladrido* “to hear a bark” are instances of pure selection. *Oír la música* “to hear music” in (22a) is also a legitimate combination because *música* “music”, as a hyponym of [sound], inherits its semantic type and may be a complement of *oír* “to hear”. This is a case of accommodation.

However, this mechanism, which founds lexical inheritance on belonging to a higher type (the hypernym), does not explain cases like *oír {el piano/el perro}* “to hear {the piano/the dog}”, because *el piano/el perro* “the piano/the dog” are not hyponyms of [sound]. If these are interpretable combinations, it is due to an operation that rescues the information contained in the QS. Here, *piano* “piano” and *perro* “dog” are

³³ This account explains in terms of pure selection a seemingly paradoxical behaviour of certain passive sentences. They require a *by*-phrase or another kind of phrase, which suggests that these phrases might be arguments. However, they are not necessarily included in the argument structure of the verb, but are rather omissible in certain contexts, or exchangeable for another, which seems to indicate that they are adjuncts. This paradox inspired Grimshaw’s proposal in 1990 about the existence of argument-adjuncts; see De Miguel (2004, 2009, 2010).

entities ([physical object] and [living being]) that emit sounds ([musical instrument] and [animal endowed with phonation device]). They are therefore words compatible with the requirement imposed by *oír* “to hear”. This additional mechanism, which demands the materialization of one possible QS meaning, is a subtype of the general coercion mechanism, which is explained below.

7.4.3. Type Coercion

This mechanism operates when a predicate imposes a certain semantic type on its argument. In its traditional formulation, coercion is a rescue mechanism for word combinations whose lexical features do not agree and should in principle be doomed to an interpretative collapse. In the classical GL definition, coercion implies lexical recategorization of an argument forced by its predicate to be semantically interpreted in a certain way without changing its syntactic type. Coercion is not activated indiscriminately or arbitrarily, but only when the underspecified definition of the “governed” word contains information that allows it to be interpreted in the sense demanded by the “governing” word (this is why *una laguna ligera* “a light lagoon”, mentioned in (3f), is unacceptable).

In the latest model revision, coercion constitutes a more general mechanism, within which two types of coercive operations can be identified. The first one is *introduction*, which coincides with what was described before as *coercion*. It is designated with this term, alluding to the predicate “covering” or “introducing” the argument in its meaning, forcing it to materialize as the type demanded. The second type is *exploitation*, which subsumes what was known as *selective binding* in the classical GL version (Pustejovsky 1995).

7.4.3.1. *Introduction*: in the first version of Pustejovsky’s model, this mechanism is generally illustrated with the example in (23d), which shows how the verb *empezar* “to begin” coerces the noun *novela* “novel”; in new terms, it is the introduction of *novela* to the verb *empezar*.

- (23) a. He empezado a trabajar en la novela
I began working on the novel
b. * He empezado la luz
* I began the light
c. He comprado la novela
I bought the novel

- d. He empezado la novela (=“he empezado a {escribirla / leerla}”)
I began the novel (=“I began {writing / reading} it”)

Empezar “to begin” is a verb that semantically selects an event in the object position (to begin doing something), as in (23a); therefore, it does not admit in principle any noun that does not denote an event, as in (23b). However, it can be constructed with nouns that in principle do not denote events, such as *novela* “novel” in (23d). To interpret this combination, Pustejovsky assumes that *to begin* imposes its selectional requirements on its complement and forces a change in its semantic type; *novel* goes from designating an [object], as in (23c), to denoting an [event]. To do this, the underspecified definition must contain information that allows for coercion. *Novel* does indeed contain this kind of information in its QS, concretely in the agentive and telic *qualia*: it is an object that exists through the event of writing and is destined to be read, as in (13a). For this reason, (23d) has two interpretations: “I began writing the novel” and “I began reading the novel”.

7.4.3.2. *Exploitation*: a mechanism that is activated when a certain component of the argument’s meaning is selected. What was called *selective binding* in the classic version of the model (Pustejovsky 1995), is a case of exploitation (materialization or binding) of a possibility contained in the word definition. This semantic operation has usually been resorted to within the GL framework to account for the polysemy of evaluative adjectives, which can be interpreted differently depending on the noun they modify. Such is the case of the adjective *excelente* “excellent” in (24):

- (24) a. Un {profesor/cuchillo} excelente
An excellent {teacher/knife}
b. Una {persona/cabellera} excelente
An excellent {person/head of hair}
c. Una comida excelente
An excellent meal

In (24a), *excelente* “excellent” refers to how teacher and knife fulfill their functions, as they are functional types specified for the telic role information. In (24b), this possibility is excluded because *persona* “person” and *cabellera* “head of hair” are natural types, which means that the adjectives are predicates of the formal or constitutive *quale* (i.e., of the external or internal qualities of the referred object). Conversely, in

(24c), *excelente* is triply ambiguous as the predicate of *comida* “meal”: *una comida excelente* “an excellent meal” may mean that “it is very good because its properties are agreeable to the palate,” that “it is of exquisite quality, created with the best ingredients” or that “it is very good for avoiding cholesterol or to guarantee good digestion”, which means that the adjective may exploit the information contained respectively in the constitutive, agentive, or telic *qualia*. There is a fourth possibility, where *comida* does not designate [food] but rather [event], because it is a complex type with two pieces of information in its formal *quale*, as in (18a); in this case, the adjective predicates the pleasantness of the eating event.

The selective binding mechanism accounts for the polysemy of the adjective *excelente* “excellent” by assuming that it binds, selects or exploits one or more pieces of the information contained in QS of the modified nouns. Similarly, the specific meaning of the polysemous *ligera* “light” in (3a, c) depends on which *quale* the adjective has bound or modified: for nouns denoting instruments (including suitcase, which is a container), the telic *quale* is modified; for those implying no particular purpose (*i.e.*, simple types, like *brisa* “breeze”), the formal or constitutive role is modified (or another feature, if it is introduced through coercion).

To sum up, adjectival polysemy is accounted for in the GL by an interpretive mechanism of a strictly lexical nature, which avoids resorting to world knowledge or postulating multiple definitions for the adjectival modifiers depending on the modified noun.

The lexical polysemy in (2b), *la clase no favorece la discusión* “the class does not favor discussion”, may now be explained. If *clase* “class” is interpreted as a physical space, its formal *quale* defines it as a [container] destined for something: specifically, to carry out an [event] in it (telic *quale*). As a container, in addition to having capacity, it has [content]: the group that attends the event that is held in the physical space (constitutive *quale*). Each interpretation may be linked to or exploited by the predicate *favorecer la discusión* “to favor discussion”, leading to one or more meanings in the combination. Lexical polysemy may thus be explained in terms similar to the logical polysemy of (2a). The same analysis can be applied to the so-called semantic polysemy, as in (2d): the word *cuadro* “painting” is a complex type, [physical object] [artistic creation] and [content]; depending on which information is exploited, the subject *de Inés* “of Inés” may be interpreted as the possessor of the physical object, the agent of creation, or the theme argument portrayed as its content. The assignment of thematic roles is thus an

epiphenomenon of the relationships between different sub-lexical features of the words” QS.

7.4.4. Co-composition

Unlike the mechanisms analyzed to this point that operate when a predicate requires a certain semantic type as its argument, co-composition is applied when it is the argument that modifies the semantics of the predicate. It is called *co-composition*, and it is activated when the argument type determines the meaning of the predicate. It explains the interpretive differences in the examples of (25):

- (25) a. Bake {a potato, garlic} _[change of state predicate]
 b. Bake {a cake, a cookie} _[creation predicate]

The predicate *to bake* is interpreted in (25a) as a change of state predicate, with the meaning “a way of cooking” (alongside {*grilling/boiling/frying*}) when it is combined with objects like *potato* or *garlic*. Conversely, it is interpreted as a creation verb in (25b) when it is combined with *a cake* or *a cookie*. This difference stems from the fact that the complements of the predicate in (25b) are non-pre-existing entities but are created precisely through the event of being baked. The co-composition of *bake* and *cake* or *cookie* leads to the sense of creation due to the identity of values in the agentive *quale* of the verb and both nouns. Because *potato* and *garlic* lack this information (potato and garlic do not come into existence in the oven), the co-composition triggers another meaning in (25a), that of a change of state from “raw” to “roasted.” This mechanism accounts for the different interpretations of the verb-noun combinations in (5) and in (3b, d) and the apparently irregular behavior of the light verbs, which in certain contexts behave as semantically “heavy” verbs (cf. (4b) vs. (4a)).

Co-composition implies interaction between the sub-lexical information of the argument and the predicate (not a restriction imposed by the predicate) in the generation of new word senses. This mechanism, combined with the other ones described above, allows dispensing with multiple definitions for the multiple meanings that words acquire in context and also capturing the systematic relationships they form. If proven adequate, GL assumptions can facilitate the study of language acquisition and processing. They may also have interesting consequences in the field of L2 teaching, as briefly shown below.

7.5 Possible Applications of the Model to L2 Teaching

The previous sections have presented a model of lexicon, the GL, which assumes that the lexicon has a highly structured and analyzable organization. This work brings up the general and universal nature of the principles and mechanisms that govern word combinations, regardless of the language to which they belong. If there is such a thing as minimal meaning elements and it is possible to derive multiple word senses in a context, we may leave behind the question of how a speaker stores infinite meanings in the mental lexicon. Instead, we can focus on establishing underspecified definitions and on uncovering the general, likely universal, principles and mechanisms that operate on them. From this perspective, the problem of language acquisition and processing becomes more approachable.

Pustejovsky defends this view, insisting on the rejection of enumerative lexicons (those that list possible word meanings) in favor of a generative lexicon that provides resources for creating different word senses. When conceptualized in this manner, the lexicon does not constitute a static repository of word senses, but rather a dynamic, flexible, and context-sensitive component (which is a promising perspective for analyzing lexically motivated grammatical phenomena). A semantic representation structured in different levels combined with generative mechanisms that integrate different word senses in a single meta-entry allows reducing the size of the mental lexicon. This explains its success in the field of computational linguistics.

Indeed, the GL framework has potential for being applied in different fields of experimental language study. It can facilitate the work of the lexicographer; actually, its assumptions underlie in part the innovative lexicographical project by Ignacio Bosque (*Diccionario combinatorio del español contemporáneo* “Combinatorial Contemporary Spanish Dictionary”, REDES, 2004). Computational linguistics has also applied it throughout the last decade, as a base for developing natural language processing systems that facilitate access to large lexical databases and in automatic translation projects. It can certainly be of help for L2 teaching, too, inasmuch as it provides tools to approach from a novel perspective the problems that classic manuals and traditional dictionaries cannot solve.³⁴

³⁴ Although this approach can make easier the work of an L2 teacher, at first sight it may seem exactly the opposite: the change of theoretical premises implies a change in both teaching and learning methodology. The learner would have to get involved in new kinds of cognitive tasks that are directed towards reflection rather than memorization.

This work concludes with some examples of errors, made by non-native Spanish speakers, that could be avoided by using an explanatory model like the one presented above.³⁵ The previously mentioned examples (1c, d) are now renumbered as (26c, d):

- (26) a. *Esperando que lea estas pocas **rayas**, le agradezco la atención y le saludo cordialmente
Hoping that you read these few **stripes**, I thank you for your attention and I send you warm greetings
- b. *Mi amiga viaja mucho y me encantan los **cuentos** de sus viajes
My friend travels a lot and I love her **short stories** about her trips.
- c. * Me he comprado unas gafas **morenas**
I have bought **brown (=dark-skinned)** sunglasses
- d. * Mis profesores de español son muy **bonitos**
My (male) Spanish teachers are very **pretty**

The hypothesis that words have an internal structure whose features must agree when two words are combined explains why the sentences in (26) are unacceptable. Traditional dictionaries do not always capture these distinctions, as in the definition of *línea/raya* “line, stripe” in the dictionary *DRAE*. This dictionary does not point out that *línea*, while a physical object (a geometric figure), is also [container] and [content], which is why *la línea está llena* (the line is full) is good. This is not a possibility for *raya* (**la raya está llena*), because it is only a physical object. The use of *raya* instead of *línea* in (26a) yields an error because, as we just saw, these words are not synonymous.³⁶

In (26b), the sentence fails because the word *cuento* “short story, tale” is specified as [+ fiction], unlike *historia* “story”, which may be marked positively or negatively for this feature. In this case, the sentence meaning requires the use of *historias* because of its non-fictional content. The use of *morenas* “dark-skinned or dark-haired” is inadequate in (26c) because this adjective must predicate of nouns denoting objects that have QS

³⁵ The data are from a Master’s thesis and a PhD thesis in preparation by Sanz (2009, 2011), an L2 teacher who has compiled errors made by students from different backgrounds and levels (from basic to advanced).

³⁶ This error was made by an Italian student, whose mother tongue encodes differently both meanings: *riga* “raya” refers both to content and physical object while *línea* only to physical object (as a matter of fact, the Italian equivalent of Spanish *subrayar* “underline” is *sottolineare*).

information relating to [pigmentation]. This is not the case of the physical object referred to by the noun *gafas*, which should be modified by *oscuras* “dark”.³⁷ Finally, the sub-lexical features of *bonito* “pretty” make it incompatible with *profesores de español* “Spanish teachers”, which can be modified by *atractivos* or *guapos* (“attractive” or “handsome”) instead.³⁸

The study of words in these combinations in terms of QS will undoubtedly prevent errors like those in (26), whether they constitute transfers from the mother tongue or are independent from the L1 of the student. The example shown in (1a) usually poses a problem for Romanian students of Spanish, because Romanian has different verbs for “to study” and “to give a class”; thus, they interpret Spanish *dar clase* only with the second meaning.³⁹ The lexical norm violations in examples (1e, f), which were uttered by native speakers, can be given a similar account: they constitute cases of feature agreement violation and therefore yield unacceptable combinations.

To sum up, I believe that the GL assumptions may help overcome some of the shortcomings of the traditional system of teaching vocabulary using the (as a rule discouraging) lists. These premises can make easier the task of correcting the students by giving them explanations based on sub-lexical features of the words to account for adequate and inadequate uses. Adopting this focus also contributes to a better understanding of the problems that students of a given L2 face while learning its lexicon and helps detect or predict them (and try to avoid them) more effectively.

References

- Anderson, Stephen R. 1990: *La fonología en el siglo XX. Reglas y representaciones*. Madrid: Visor. (*Phonology in the twentieth century*. University of Chicago Press. 1985).
- Bartra, Anna. 2009. La enseñanza del léxico. In De Miguel, Elena (ed.), *Panorama de la Lexicología*, 435-463. Barcelona: Ariel.

³⁷ Both errors were made by native English speakers. In their mother tongue, *story* and *brown* are not specified (either positively or negatively) for the features [±fiction] or [±pigmentation], respectively.

³⁸ This may also be an erroneous translation of the English word *nice*. Instead of being predicated of the physical appearance (formal *quale*) it may be predicated of the teacher’s personality, of being accessible or friendly (this information could be encoded in the telic *quale*, available for the word *profesor*). [Observation of Olga Batiukova, p.c.]

³⁹ I thank Gabriela Draghici for this observation.

- Battaner, Paz. 2009. Análisis lexicográfico del verbo *hacer*: el verbo *hacer* en un diccionario de aprendizaje. In Juan Manuel García Platero and María Auxiliadora Castillo Carballo (eds.) *Investigación lexicográfica para la enseñanza de lenguas*, 69-89. Málaga: Universidad de Málaga.
- Batiukova, Olga. 2008. Aplicaciones lexicográficas de la teoría del Lexicón Generativo. In Elena De Miguel, Santiago U. Sánchez, Ana Serradilla, Romana Radulescu and Olga Batiukova (eds.), *Fronteras de un diccionario. Las palabras en movimiento*, 231-268. San Millán de la Cogolla: Cilengua.
- Bosque, Ignacio. 1989. *Las categorías gramaticales. Relaciones y diferencias*, Madrid: Síntesis.
- . 2001. On the weight of light verb predicates. In Julia Herschenson, Karen Zagona and Enrique Mallén, (eds.), *Features and interfaces in Romance*, 23-38. Amsterdam: John Benjamins.
- . (2004). Combinatoria y significación. Algunas reflexiones. In *REDES. Diccionario combinatorio del español contemporáneo*, LXXVII-CLXXIV, Madrid: SM.
- . (dir.). 2004. *REDES. Diccionario combinatorio del español contemporáneo*, Madrid: SM.
- Cohen, L. Jonathan. 1986. How is conceptual innovation possible? *Erkenntnis* 25: 221-238.
- Cruse, David A. 2004. *Lexical semantics*. Cambridge: Cambridge University Press.
- De Miguel, Elena. 2004. La formación de pasivas en español. Análisis en términos de la estructura de *qualia* y la estructura eventiva. *Verba Hispanica* XII: 107-129.
- . 2006. Cosas que ocurren dentro de las palabras y entre palabras: datos e hipótesis acerca de la estructura del léxico y de ciertas estructuras léxicas. In Elena De Miguel, Azucena Palacios and Ana Serradilla (eds.), *Estructuras léxicas y estructura del léxico*, 7-31. Frankfurt am Main: Peter Lang.
- . 2009. La Teoría del Lexicón Generativo. In Elena De Miguel (ed.), *Panorama de la Lexicología*, 337-368. Barcelona: Ariel.
- . 2010. La polisemia de los verbos soporte. Propuesta de definición mínima. In Paz Battaner and Elisenda Bernal (eds.), *Los verbos en los diccionarios*. Barcelona: Edicions de l'IULA, Universitat Pompeu Fabra.
- . 2011. En qué consiste ser verbo de apoyo. In María Victoria Escandell Vidal, Manuel Leonetti and Cristina Sánchez (eds.), *60 problemas de gramática, dedicados a Ignacio Bosque*, 139-146. Madrid: Akal.

- De Miguel, Elena and Marina Fernández Lagunilla. 2000. El operador aspectual *se*. *Revista Española de Lingüística* 30. 1: 13-43.
- De Miguel, Elena and Marina Fernández Lagunilla. 2006. La naturaleza léxica del aspecto composicional. In *Actas del VI Congreso de Lingüística General (2004)*, II (I): 1767-1778. Madrid: Arco/Libros-Universidad de Santiago de Compostela.
- García Platero, Juan Manuel and María Auxiliadora Castillo Carballo (eds.). 2009. *Investigación lexicográfica para la enseñanza de lenguas*. Málaga: Universidad de Málaga.
- Geeraerts, Donna. 2002. The theoretical and descriptive development of lexical syntax. In Leila Behrens and Dietmar Zaefferer (eds.), *The lexicon in focus. Competition and convergence in current lexicology*, 23-42. Frankfurt am Main: Peter Lang.
- González Cobas, Jacinto. 2010. Diccionarios y rasgos de definición: los nombres de instrumentos de medida. In *Actas del XXXIX Simposio Internacional de la Sociedad Española de Lingüística*. Santiago de Compostela, CD-Rom version.
- Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: The MIT Press.
- Jackendoff, Ray S. 2002. *Foundations of Language. Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Katz, Jerrold J. 1972. *Semantic theory*. New York: Harper and Row.
- Lipka, L. 1990. *An outline of English lexicology: Lexical structure, word semantics, and word-formation*. Tuebingen: Max Niemeyer.
- Luque Durán, Juan de D. 2004. Aspectos universales y particulares del léxico de las lenguas del mundo. *Estudios de Lingüística del Español (ELiEs)* Vol. 21 (<http://elies.redris.es/elies21>).
- Mairal, Ricardo and Francisco Cortés. 2009. Modelos funcionales. In Elena De Miguel (ed.), *Panorama de la lexicología*, 247-279. Barcelona: Ariel.
- Mendikoetxea, Amaya. 2006. En busca de los primitivos léxicos y su realización sintáctica: del léxico a la sintaxis y viceversa. In Cabré, Teresa (ed.), *Lingüística Teòrica: Anàlisi i Perspectives II*, *Catalan Journal of Linguistics. Monografies*: 55-102. Bellaterra: Servei de Publicacions de la UAB.
- . 2009. Modelos formales. In Elena De Miguel (ed.), *Panorama de la lexicología*, 331-335. Barcelona: Ariel.
- Moravcsik, Julius M. 1975. *Aitia* as generative factor in Aristotle's philosophy. *Dialogue (Canadian Philosophical Review)* 14. 4: 622-638.

- . 1991. What makes reality intelligible? Reflections on Aristotle's theory of *Aitia*. In L. Judson (ed.), *Aristotle's physics*, 31-47. Oxford: Clarendon Press.
- Moreno Cabrera, Juan Carlos. 1997. *Introducción a la lingüística. Enfoque tipológico y universalista*, Madrid: Síntesis.
- . 2003. *Semántica y gramática. Sucesos, papeles semánticos y relaciones sintácticas*, Madrid: Antonio Machado.
- Murphy, M. Lynne 2010. *Lexical meaning*. Cambridge: Cambridge University Press.
- Peeters, Berts. 2000. Setting the scene: Some recent milestones in the lexicon-encyclopedia debate. In Berts Peeters (ed.), *The Lexicon-encyclopedia interface*, 1-52. Oxford: Elsevier.
- Pérez Saldanya, Manuel and Llorenç Comajoan. 2011. *El tiempo y el aspecto verbales en el español como segunda lengua*. Unpublished ms.
- Pustejovsky, James. 1988. The geometry of events. In Carol Tenny (ed.), *Studies in generative approaches to aspect. Lexicon Project Working Papers* 24, 19-39. Cambridge, MA: The MIT Press.
- . 1991. The Syntax of Event Structure. In Beth Levin and Steven Pinker (eds.), *Lexical and conceptual structure*, 47-81. Oxford: Blackwell.
- . 1993. Type coercion and lexical selection. In James Pustejovsky (ed.), *Semantics and the lexicon*, 73-94. Dordrecht: Kluwer.
- . 1995. *The Generative Lexicon*. Cambridge, MA: The MIT Press.
- . 2006. Type theory and lexical decomposition. *Journal of Cognitive Science* 6: 39-76.
- . 2008. From concepts to meaning: The role of lexical knowledge. Talk given at the 18th International Congress of Linguistics, July 21-26, Seoul, Korea.
- Pustejovsky, James and Bran Boguraev. 1993. Lexical knowledge representation and natural language processing. *Artificial Intelligence* 63: 193-223.
- . 1994. A richer characterization of dictionary entries: The role of knowledge representation. In Beryl T.S. Atkins and Antonio Zampolli (eds.), *Computational approaches to the lexicon*, 295-311. Oxford: Clarendon.
- . 1996. The problem of lexical ambiguity. In James Pustejovsky and Bran Boguraev (eds.), *Lexical semantics. The problem of polysemy*, 1-14. Oxford: Clarendon.
- Pustejovsky, James and Pierrette Bouillon. 1996. Aspectual coercion and logical polysemy. In James Pustejovsky and Bran Boguraev (eds.), *Lexical semantics. The problem of polysemy*, 133-162. Oxford: Clarendon.

- Pustejovsky, James and Federica Busa. 1995. Unaccusativity and event composition. In Pier Marco Bertinetto, *et al.* (eds.), *Temporal reference, aspect and actionality* Vol. 2, 159-177. Turín: Rosenberg and Sellier.
- Pustejovsky, James and Anna Rumshisky. 2008. Between chaos and structure: Interpreting lexical data through a theoretical lens. In Patrick Hanks (ed.), *Special issue of the international journal of lexicography on the legacy of John Sinclair*, 337-355. Oxford: Oxford University Press.
- . 2009. emEval-2010 Task 7: Argument selection and coercion. In *NAACL 2009 Workshop on semantic evaluations: Recent achievements and future directions (SEW-2009)*. Boulder, Colorado.
- RAE. 2001. *Diccionario de la lengua española*. Madrid: Espasa.
- Sanz, Ana. 2009. *El error de léxico en la enseñanza de ELE y la información subléxica de las palabras*, Postgraduate research report. Madrid: Universidad Autónoma de Madrid.
- . 2011. *El error de léxico en la enseñanza de ELE: un análisis subléxico*, Unpublished ms. Universidad Autónoma de Madrid.
- Singleton, David. 2000. *Language and the lexicon. An introduction*. London: Arnold.
- Val Álvaro, José Francisco and M.Carmen Horno Chéliz. 2010. *La gramática del sentido: Léxico y sintaxis en la encrucijada*. Zaragoza: Prensas Universitarias de Zaragoza.
- Van Valin, Robert D. 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.

