(FORTHCOMING) PIUS TEN HACKEN (ORG.) "MEANING AND LEXICALIZATION OF WORD FORMATION"

SEMANTIC COINDEXATION: EVIDENCE FROM PORTUGUESE DERIVATION AND COMPOUNDING

ALEXANDRA SOARES RODRIGUES AND GRAÇA RIO-TORTO[1]

1. INTRODUCTION

The parallel between meaning construction in derivation and compounding has received little discussion. Some works such as Lieber (2004) and Fradin (2005) focus on it. However, more empirical data is needed to contribute to the understanding of how meaning construction works. Our contribution brings more data on the phenomena by comparing derivation and compounding in Portuguese.

The aim of this article is to analyse the way meaning construction occurs in derivation and compounding. We try to answer the following questions:

1. How do words formed by derivation and compounding get their meaning?
2. What are the factors involved and what is the balance between them?
3. Are compounding and derivation rules sensitive to the semantic of their bases in the same manner?

1 The article is the result of close collaboration between the authors. For academic purposes, Rodrigues (Instituto Politécnico de Bragança and CELGA) is responsible for section 2. and Rio-Torto (Universidade de Coimbra and CELGA) for section 3. The remaining text is subscribed by both authors.

The analysis is focused on Portuguese deverbal nouns and adjectives (section 2) as well as on nominal compounds (section 3) formed by noun-noun $[NN]_N$ and noun-adjective $[NA]_N$.

We believe that the formation of meaning in the word is independent of syntax (cf. Lieber and Scalise (2007) regarding compounds), since there is a discrepancy between the meanings provided by syntactic arrangement and argument structure and the developed meanings of the coined word.

We state that semantic coindexation is the responsible for the meaning construction in word formation. On our proposal, coindexation operates between semantic features of the constituents (affix and base or compound bases) and those of the 'maximal semantic frame' (Fillmore 1977; Langacker 1987; Jackendoff 1997, 2002) associated with them.

Coindexation is ruled by what we will call the maximal compatibility principle.According to our proposal, semantic coindexation between the involved features is dependent on the degree of semantic similarity between them. [2] This principle prevents chaotic linking between features, because it only allows the linkage of those that best fit semantically with each other.

The semantic framework adopted here corresponds to a conceptualist version of semantic processing (Jackendoff 2002, 2007), according to which the meanings of words must conform to human categorisation, to mental representations connected to perception and action and to the speaker's experience with language and the world. Thus**,** the meaning of complex words come from the lexical units involved, as well as from other information sources, such as referential and/or pragmatic (Jackendoff 1997, 2002, 2007).

2 The concept of coindexation that we use here is different from Lieber's concept (Lieber 2004). Contrarily to her, we propose that coindexation is strictly semantical and does not work with syntactical or argumental elements. The concept of compatibility is also different from the one proposed by Langacker (1987), since his one admits sentences and proprieties of whatever nature.

2. DERIVATION

2.1 Absence of syntactic factors in meaning construction in word formation

Traditionally, deverbal derivations are seen as the result of either the projection of the argument structure of the verbal base, in the case of event deverbal nouns (e.g. Grimshaw 1990), or of one of the arguments of the base verb, in the case of agent deverbal nouns and adjectives (Rappaport Hovav and Levin (1992)). This perspective has already been questioned by authors such as Hoekstra and Putten (1988) and Rodrigues (2008).

As an example, let us consider agent nouns/adjectives such as *colonizador* (‘colonizer’). In this case, apparently, the argument structure of the verbal base creates the agent meaning of the noun. Without giving many details, the external argument of the verb *colonizar* (‘to colonize’) is topicalized in the meaning of *colonizador*. The internal argument of *colonizar* is still available outside the noun (cf. 1).

(1) O colonizador do Peru.
The colonizer of.the Peru
‘The colonizer of Peru.’

The problem arises when we apply this syntactic explanation to nouns and adjectives such as the following:

(2) a. *lambedor*$_{N}$ (from *lamber* ‘to lick’), which, apart from the meaning of ‘licker’, also has the meaning of ‘syrup’;
b. *chovedor*$_{A}$ ‘that makes rain’;

c. *suador*$_{A/N}$ ‘that makes sweat’, besides the more prototypical meaning of ‘that sweats’.

The syntactic explanation does not fit with these deverbal nouns. Let us observe the syntactic, the argument and the lexical-conceptual structures of their verbal bases:

(3) a. *Chover* ‘to rain’: intransitive verb. It has an expletive subject that corresponds to a syntactic function that is argumentally empty (no theta-role);

b. *Suar* ‘to sweat’: Intransitive verb (unergative): external argument (internal cause [– volitional]);

c. *Lamber* ‘to lick’: Transitive verb: external argument (external cause [+ volitional]); internal argument (theme).

Let us compare these data with the structures of the nouns/ adjectives (table 1):

Verb	Noun/Adjective
Chover: *Chove*. Rains (It rains.)	*Chovedor:* *Substância chovedora* Substance that.makes.rain.FEM.
Suar: *O João suou*. The John sweated. ‘John sweated.’	*Suador*: *Exercício suador* exercice that.makes.sweat
Lamber: *O João lambeu o xarope*. The John licked the syrup. ‘John licked the syrup.’	*Lambedor*: *O João comprou um lambedor*. The John bought a ∅ that.makes.lick ‘John bought a syrup.’

Table 1. Comparison, in context, between the argument and the lexical-conceptual structures of verbs and their deverbal nouns.

From the data shown in table 1, we can see that *chovedor*, *suador* and *lambedor* display an agent meaning ‘external cause’. However, their verbal bases lack the argument (external cause) that would correspond to the one that appears in the

argument structure of the nouns. Where does this 'external cause' meaning come from?

In these cases, the 'external cause' meaning must come from a maximal semantic frame, which is not particular to any lexeme, but available in the conceptual structure in general.[3] This maximal semantic frame also explains deverbal nouns with meanings such as 'locative' that do not correspond to any argument of the argument structure of the verb base. This is the case of *bramadeiro* 'place where deer join together when in rut', from *bramar* 'to bellow (deer)'; *miradouro* 'viewpoint', from the verb *mirar* 'to watch'; and *matadouro* 'slaughterhouse', from the verb *matar* 'to kill'. None of these verbs display a locative argument in their argument structure.

These are not marginal examples of deverbal nouns. In fact, many of them do not correspond to the argument structure of the base in what concerns meaning and in what concerns their proper argument structure capacity (Rodrigues 2008: 80-93). These examples show that the relation between the meaning of the deverbal nouns and the verbal base must be founded on fine-grained semantic structures.

2.2. Semantic coindexation

We propose that meaning construction in word formation is sustained by the combination of semantic features of the base, the affix (if there is an affix involved, as is the case for the lexemes under analysis)[4], and the maximal semantic frame. The mechanism that is responsible for the activation of those connections is coindexation. In contrast to Lieber (2004), we propose that coindexation is a purely semantic mechanism (Rodrigues 2008: 60-69). The reason we put aside any kind of syntactic

[3] This 'maximal semantic frame' refers to a conceptual universe related to each lexeme (Langacker 1987; Pustejovsky 1995; Jackendoff 2002).

[4] In the case of converted deverbal nouns, which lack affixation, the parameters involved in their meaning formation come from the verbal base and the maximal semantic frame.

intervention in coindexation has to do with the fact that there may be no relation between the arguments of the verb and the meaning of the deverbal noun, as we have seen in section 2.1.

To understand coindexation, we must conceive of semantics as a domain structured in tiers (Jackendoff 2002). The components of these tiers of a lexeme are able to be dynamically linked to components of other tiers or of the same tiers of other lexemes (Rodrigues 2008: 60). It may be objected that a process of coindexation totally based on semantic structures would lead to an overgeneration of derivations. However, semantic coindexation is based on the degree of compatibility between the features of the base, the features of the affix and the features of the maximal semantic frame (Rodrigues 2008: 227-274).

Let us give an example: the affix *-dor* prototipically generates agent nouns. In Rodrigues (2008: 340-353), we have explained why this occurs. It is because the semantic feature of *-dor* is [that has the function of]. The meaning of this feature is very close to the meaning of the feature [agent] of the lexical-conceptual structure of a verb, but it is not so close, for instance, to the feature [place]. Indeed, [that has the function of] alludes to something or someone that will actualise a given event. This is a meaning quite similar to [agent]. Semantically, due to the presence of an [active] feature, the degree of compatibility between [that has the function of] and [agent] is higher than the one between [that has the function of] and [place].

We have observed in Rodrigues (2008) that the most prototypical meanings and derivatives result from the coindexation of features that are semantically closer to each other. If the feature [that has the function of] of *-dor* coindexes with the agent feature of the lexical-semantic structure of the verbal base, the obtained meaning will be ‘agent’, which is a prototypical meaning of *-dor* derivatives. This situation

illustrates a process of maximal compatibility between the features of the suffix and the base. If the same feature coindexes with [place], the obtained meaning will be 'place', which is far from being a prototypical meaning of *-dor* nouns.

Due to this need for compatibility between the features, an overgeneration of derivations is avoided. The maximal compatibility between the features of the affix and those of the base represent the most prototypical derivatives/meanings of that paradigm (*e.g.*, *-dor* nouns meaning 'agent', such as *conquistador* 'conqueror', *avaliador* 'evaluator'). A minimal compatibility between the features leads to the least representative derivatives of that paradigm (*e.g. -dor* nouns meaning 'place', such as *toucador* 'dressing table'). The frontiers provided by *minimal* and *maximal* are dependent on the kinds of meanings (lest and most prototypical) that the derivatives of each suffix display.

Following authors such as Plag (1999, 2002), we assume that affixes are provided with semantic features. An affix is not simply a formal operator of a certain word formation rule. This explains some kinds of constraint holding between affix and base; that is, why an affix occurs with one kind of verbal base (*e.g. comiseração*$_{N}$ from *comiserar*$_{V}$ 'to move to pity') and not with other kinds of verbal base (*e.g.* **envelheceção*$_{N}$ from *envelhecer*$_{V}$ 'to become old') (cf. *envelhecimento* vs. **comiseramento*), since both affixes generate event deverbal nouns.

Semantic features of the affix are not accessible when the affix is on its own. As a non-autonomous morpheme, semantic structures of the affix are only observable when the affix is in the whole of the derived noun it had generated. This is to say that the semantic charge of the affix is placed in an implicit structure. We need to compare the derivatives of that affix between them and with the derivatives of other affixes that operate in the same rule to determine it.

As an example, consider event deverbal nouns. Although sharing the same verbal base and a general meaning 'event of V', the nouns reveal different semantics according to the affix. Let us compare event nouns from the verb *andar* 'to walk'. The noun with the suffix *-nça* [[*anda*]$_{V}$*nça*]$_{\mathbf{N}}$ means 'adventure, journey'; the noun with the suffix *-mento* [[*anda*]$_{V}$*mento*]$_{N}$ means 'speed or way of something going'; and the noun with the suffix *-dura* [[*anda*]$_{V}$*dura*]$_{N}$ 'physical way of moving'. The differences between their meanings come from the coindexation of semantic features of each affix with semantic features of the base.

The semantic features of the base belong to the event structure and to the lexical-conceptual structure. We are not going to present here all the features of verbal bases and affixes delimited in Rodrigues (2008: 227-274) from the analysis of 8414 deverbal nouns constructed with 23 affixes. Let us consider only the event structure. We have determined the following event features available in the verbal bases:

- – [punctual] – the event occurs at a particular point on the temporal line and not along that line (e.g. *estalar* 'to click');
- – [durative] – the event occurs along the temporal line (*viver* 'to live');
- – [composed of individuals] – This feature was presented by Lieber (2004: 136). It refers to an event that is symmetrically divided into parts repeated along the temporal line (*saltitar* 'to hop').
- – [composed of different operations] – the event is composed of different subevents (*e. g. conduzir* 'to drive (a car)' is composed of subevents such as 'clutching', 'braking', 'changing direction with the steering wheel', etc.).
- – [point of departure] – the event has a beginning. This feature permits us to distinguish durative verbs such as *distar* 'to be distant from', which is a state

verb, from durative verbs such as *caminhar* ‘to walk’. *Distar* does not indicate an event with a clear temporal beginning. On the contrary, *caminhar* has an implicit beginning. This difference shows that event structure contains subcomponents and does not behave like an indivisible whole.

- – [point of arrival] – the event has an end point. This allows us to distinguish a verb such as *construir* ‘to build’ from a verb such as *trabalhar* ‘to work’. The first one has a point of arrival, whilst the second one does not.
- – [telic] – a telic verb presupposes that a change of state occurs. For instance, *estar* ‘to be’ vs. *cozinhar* ‘to cook’. The second one is [+ telic], whilst the former is not.
- – [perfect] – the event is irreversible and not prolongable (e.g. *matar* ‘to kill’).

These features do not correspond to classes; i.e., each feature on its own does not characterize the event type of the verb. Each verb may have a set of features. At the beginning of our study, we tried to use those classes. However, it has been shown to be useless, since semantic coindexation does not operate with semantic boxes, but with subcomponents of those boxes.We mean by this that we have tryed to check if the selection of each suffix to the formation of deverbal nouns was dependent on the event class of the verbal base (*e. g.* Vendler’s classes: *accomplishment*, *achievement*, *activity*, *state*). What we have observed was that there was no relation between the event class of the verbal base and the selected suffix. For instance, some verbs of accomplishment such as *enrolar* (‘to wrap’) select the affix *-mento* (*enrolamento* ‘event of V) and so do some verbs of achievement such as *salvar* (‘to save’) (*salvamento* ‘event of V’), some verbs of activity such as *respirar* (‘to breathe’) (*respiramento* ‘event of V’) and even some verbs of state such as *preceder* (‘to

precede') (*precedimento* 'event of V'). However, not all verbs of those classes select this suffix.

We have needed to observe the features that characterised each event structure of each verb. Those features, and not the entire event structure as a whole, have revealed to be important not only to the selection of the suffix, but also to the determination of the meaning of the deverbal noun (Rodrigues 2008: 201-202).

The verbs *acalmar* 'to calm down' and *relaxar* 'to relax' contain features such as [durative], [telic] and [point of arrival], since they behave as accomplishment verbs. The event deverbal nouns from these verbs are *acalmamento*, *acalmação* and *relaxamento* e *relaxação*, respectively. Although being event deverbal nouns, *acalmamento* and *relaxamento* present different semantic shades in comparison with *acalmação* and *relaxação*. Nouns with the affix *-mento* display a meaning of 'state' that co-occurs with the course of the event. Nouns with the affix *-ção* have a meaning of 'state' that does not co-occur with the course of the event; it occurs after or as a consequence of the point of arrival of the event.

This can be explained if we assume that affixes have a semantic structure. The affix *-mento* contains the feature [process], whilst *-ção* is characterized by the feature [effectuation]. Notice that these features are not included in the above list, since that list only shows semantic features of the verbal bases, not semantic features of the affixes. We are not going to show here all the semantic features of the affixes that we have delimited in Rodrigues (2008: 227-274). We will stick to the affixes shown here.

The feature [effectuation] refers to an event that is offered as actualized and completed. Contrary to [effectuation], the feature [process] refers to the course of the event and not its ending. It stresses the proceeding of the event and not its conclusion. Once again, we need to emphasise that semantic features of affixes become available

to our explicit knowledge when we compare deverbal nouns with different affixes and the same verbal bases between them and deverbal nouns with the same affix and different verbal bases.

Producing *relaxamento* and *calmamento* uses the following procedures: the affix *-mento* has the feature [process]. This feature is maximally compatible with the feature [durative] and minimally compatible with the feature [point of arrival]. Thus, *-mento* or, specifically, its feature, will coindex with [durative] and not with [point of arrival]. In consequence, *relaxamento* and *calmamento* mean a state that co-occurs with the course of the event. On the contrary, *-ção* has the feature [effectuation]. This feature is maximally compatible with the feature [point of arrival] and minimally compatible with the feature [durative]. Thus, *acalmação* and *relaxação* mean a state that occurs after the end of the process, that is, at the point of arrival (Rodrigues 2008: 291-315).

This difference between the semantic features of *-ção* and *-mento* explains why *-mento* attaches to durative verbs such as *balancear* 'to swing', *deslizar* 'to slide', *enramar* 'to develop branches (in a tree)', *espigar* 'to ear (cereal)'. The feature [process] of this affix prefers the feature [durative] of these verbs. We find the nouns *balanceamento*, *deslizamento*, *enramamento*, *espigamento*, but we do not find **balanceação*, **deslização*, **enramação*, **espigação*.

From this perspective, the maximal semantic compatibility principle explains why *-mento* attaches to *-ec-* and *-esc-* verbs and not to *-iz-* and *-ific-* verbs. Nouns such as *envelhecimento* 'growing old', from *envelhecer* 'to grow old', *amarelecimento* 'yellowing' from *amarelecer* 'to yellow', *robustecimento* 'strengthening' from *robustecer* 'to strengthen' are common. Nouns of these kinds of verbs with *-ção* turn out to be ungrammatical (**envelheceção*; **amareleceção*;

**robusteceção*). This is because *-mento* selects verbs where the feature [durative] is foregrounded.

The same factor explains why *-mento* is the affix that occurs with verbs such as *abairrar* 'to divide into wards' (*abairramento*) and *arruar* 'to divide into streets', which indicate the unbounded division of an object in infinite parts.

Apart from these differences in the meaning of 'event' and the correlated consequences in the constraints between bases and affixes, semantic coindexation also reveals itself in the concrete meanings that the derivatives of each affix exhibit.

Let us observe the following deverbal nouns in table 2:

Verbal bases	*affix -dura*	*affix -ção*	*affix -mento*
serrar 'to saw up'	*serradura* 'sawdust'	*serração* 'event of V; sawmill'	
cevar 'to feed on'	*cevadura* 'rests of the bird that a bird of prey has been fed on'	*cevagem* 'the action of feeding' (causative construction)'	
amolgar 'to dent'	*amolgadura* 'dent, depression'		*amolgamento* 'event of V'
entrançar 'to plait'	*entrançadura* 'plait'		*entrançamento* 'event of V
maçar 'to bruise'	*maçadura* 'bruise'		
pisar 'to bruise'	*pisadura* 'bruise'		
limar 'to file'	*limadura* 'filing (steel)'		
varrer 'to sweep'	*varredura* 'sweepings'		
abotoar 'to button'	*abotoadura* 'set of buttons'	*abotoação* 'event of buttoning'	

Table. 2 Verbal bases and their deverbal nouns with affixes -dura, -ção and -mento

All of the deverbal nouns presented in table 2 have an event meaning related to their verbal bases. However, *-dura* nouns also have concrete meanings of 'portion',

‘residue’, ‘amounts’, ‘concrete result’, which are absent from the other deverbal nouns. Where do those concrete meanings come from?

In the theory we propose here, the suffix *-dura* has the feature [referentiation]. Once again, this feature only comes to light indirectly when we observe the meanings of *-dura* nouns in comparison to the other event affixes derivatives. Special data are provided if we compare deverbal nouns of the same verb with different event affixes. What we observe in *-dura* nouns is that, besides the meaning of ‘event’, many of them manifest a concrete meaning that can be subsumed as ‘something concrete that results from the event’, ‘a physical result of the event’. This is specified as ‘residue’ (*serradura* ‘sawdust’, *cevadura* ‘remains of the bird that a bird of prey has been fed on’, *varredura* ‘sweepings’), ‘physical wound’ (*maçadura* ‘bruise’, *pisadura* ‘bruise’). This affix only attaches to verbs that possess a concrete meaning. This emphasizes the semantic compatibility between affix and base.

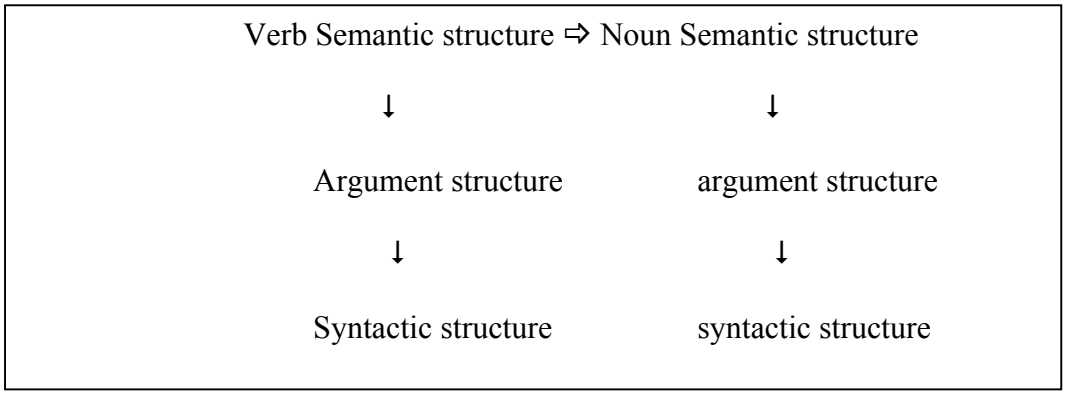
Nevertheless, what explains the construction of these particular meanings is coindexation. All the verbal bases of these nouns possess the feature [telic], among others. The feature of the affix, which is [referentiation], indicates a segmentation and an identification of a certain event, detaching it from the continuum of realia. The semantic effect of this affix is not to provide an event shade of the event, i.e., [durative], [actualized], etc., but simply to identify it as a referent (Rodrigues 2008: 315-320). Therefore, the feature [referentiation] is maximally compatible with the feature [telic] of the verbs. The segmentation and identification of the referent of [telic], operated by the feature of the affix, result in these peculiar concrete meanings.

2.3 Conclusions on derivation

In relation to meaning construction in derivation, we have come to the following conclusions:

i) meaning construction in derivation is operated in a first phase where argument structure does not intervene. In this phase, there is the intervention of lexical-conceptual structure, and of the 'maximal semantic frame' associated to each lexeme. That intervention consists in the coindexation of semantic features belonging to the lexical-conceptual structure of the base, or to a maximal semantic frame, which is subsumed in mental conceptual structures (Jackendoff 2002; Pustejovsky 1995). Meaning construction in word formation does not occur in the level of argument structure. If the deverbal noun has an argument structure, it is the result of the combination of a series of factors (among them the kind of meaning of the noun, and the affix that formed it[5]). It is not inherited from the verbal base.

This hypothesis is outlined in figure 1.



[5] This is not to say that the affix contains in itself the argument structure, in the case of event deverbal nouns, or a specific argument in the case of agent deverbal nouns. We mean that the affix, because of its meaning, functions as a constraint on the development of argument structure in the deverbal product. This explains why deverbal nouns from the same verb offers different argument structure solutions according to the affix that built them. For instance, *apagão* 'power blackout' does not have any argument structure, while *apagamento* 'switching off; putting out' does (from *apagar* 'to switch off; to put out'). This also shows that argument structure in deverbal nouns is not inherited from the base. Otherwise, every deverbal noun would have an argument structure and this is not the case. By lacking argument structure, we mean the incapacity of the deverbal noun to function as a head that would induce argument positions in the phrase. Both *apagão* and *apagamento* mean an 'event'. However, the first one does not function as a head, whilst the second one does (see example 4).

(4) a. **O apagão da luz pelo Mário.*
The switching.off of.the light by.the Mário
'The switching off of the light by Mário.'

b. *O apagamento da luz pelo Mário.*
The switching.off of.the light by.the Mário
'The switching off of the light by Mário.'

Grimshaw (1990: 45) and Grimshaw & Williams (1993: 98) distinguish both kinds of nouns by the designations 'complex event nominals' (*e. .g. apagamento*) and 'simple event nominals' (*e. g. apagão*).

Figure 1.Directional relations between different structures of verbal base and its deverbal noun

ii) the mechanism that is responsible for meaning construction is coindexation. Coindexation operates between features of the affix, of the base or of the maximal semantic frame associated with it. Although strictly based on semantic parameters, coindexation is ruled by the maximal compatibility principle. This principle prevents chaotic linking between features, because it only allows the linkage of those that best fit semantically with each other.

3 COMPOUNDS

A compound is a plurilexematic structure that is used as a holistic denomination, that resists any internal alteration and that is characterized by psycholinguistic and denotational unicity. This conception is grounded on compound investigation carried out by Lieber and Scalise (2007), Rio-Torto and Ribeiro (2009, 2010) and Scalise and Bisetto (2009) and is supported by a cross-disciplinary approach on compounding of several languages (Scalise and Vogel 2010) studied by Morbocom team (http://morbo.lingue.unibo.it/).

Portuguese $[NN]_N$ and $[NA]_N$ compounds are conceived of as morphological objects,[6] whose components are linked by a modification relationship (Rio-Torto and Ribeiro 2009). As the compounds under analysis always have the same syntactic

[6] Independently of their internal structure and of their gradual nature — from a more to a less phrasal level —, compounds are conceived as lexical constructions (Booij 2009) whose functioning is ruled by morpho-semantic — not by syntactic — principles (Lieber and Scalise 2007).

structure ($[NN]_N$ or $[NA]_N$)[7], the diversity of semantic relations they express cannot be explained syntactically, but only semantically.

The understanding of compounds is anchored on the assumption that there exists a semantic relation between the referents of the two concepts being combined (NN) or that one or more properties of the modifier constituent are attributed in some way to the head concept (NA). What are the principles and the devices that govern this semantic coindexation? How should the emergence of idiomatic meanings be explained?

On the basis of the discussion in this section, we will claim that a theory of enriched meaning and processing (Jackendoff 1997) is necessary when simple composition does not suffice.[8] Semantic procedures, like coercion, reference transfer, figurative shifts, conceived in an "enriched composition frame", are responsible for the conventionalized meaning of word.

3.1 Some assumptions

Compounding is associated with idiomaticity.[9] Indeed, there are compounds whose meanings are compositional and transparent, but many compounds have an idiomatic meaning: a *garbage man* is a 'man who handles garbage', but a *snow man* is a 'simulated man made of snow'.

We state that the meaning of a compound incorporates, even in a sophisticated, inexpectable and idiomatic fashion, the meaning of its components.

---

7 The relational (bomba atómica 'atomic bomb') or qualitative (mau feitio 'bad temper') nature of the adjective is not relevant in the cases under analysis. For the classifying and qualifying functions of Portuguese adjectives see Rio-Torto and Ribeiro (2009: 282-284).

8 Adapting Jackendoff words (1997: 49) about sentences, we suggest that the meaning of a compound «may contain, in addition to conceptual contents of its LCSs, other material that is not expressed lexically, but that must be present in conceptual structure either (i) in order to achieve well-formedness in the composition of the LCSs into conceptual structure (coercion, to use Pustejovsky's term), or (ii) in order to satisfy the pragmatics of the discourse or extralinguistc content».

9 For an overview about the relations between compositionality and idiomaticity, see Cruse: 2004: 68-77.

Nevertheless, the meaning of a compound is not necessarily confined to the meanings of its parts and of the rules by which they are combined. The maximal semantic frame of a compound includes all the features associated with it in a specific civilizational universe, namely the features associated with the profiles, the roles and/or the proper functions of denotata, as well as the pragmatic purposes words can serve.

We claim that idiomaticity is the result of forced meaning shifts regarding the compositional one (Rio-Torto and Ribeiro 2010). Compositionality and idiomaticity vary inversely. However, as empirical data illustrate it (cf. (6)), sometimes earlier stages of semantic construction may display the underlying compositionality dissipated by idiomaticity (Rio-Torto and Ribeiro 2010).

3.2 Trends of idiomaticity

Within a compound, the head and the modifier must be coindexed semantically in accordance with (i) the (maximal) information they convey and (ii) the possible grammatical and LCS relationships linking them (Bisetto and Scalise 2005). When the final meaning does not include the compositional meanings of the constituents, other semantic devices must be managed. In order to guarantee the internal compatibility and plausibility of the whole**, coercion procedures** are activated. Metaphor, metonymy, referential shift and/or specialization are often activated for this task, and they are mainly responsible for the idiomaticity of the compounds, as the following (4-6) examples illustrate.

(4) $[NN]_N$ *visita-relâmpago* (lit. visit-lightning)

(5) $[NN]_N$ *pontapé* (lit. fronting of foot)

(6) $[NA]_N$ *saco azul* (lit. bag blue)

Visita-relâmpago (lit. visit-lightning) denotes not a visit of lightning, but 'a flying visit, an unexpected and brief one': a **metaphor** is activated as the features [brief and unexpected] are transferred from the lightning to a visit. *Pontapé* (lit. the fronting of foot) denotes a kick, an aggression with the foot; a **metonymy** is activated, as the act is denominated by the 'actor [foot]'.Usually, when figurative tools are activated a reference transfer occurs, e.g. in (6). *Saco azul* (lit. blue bag) denotes not any specific blue bag, but illicit funding. The older motivation — an (ancient) bag containing money from an unofficial source and lined with blue tissue — is lost. The metonymic meaning is not semantically compositional because the meaning of the whole is not computable from the meaning of the constituents. The examples (5-6) display a *reference transfer* between the denotation/the ontological class of the head and the one of the compound (foot>physical aggression; bag>funding). Their interpretation is opaque to native speakers, since they have not previously been exposed to it.

Semantic specialization between coindexed compound members can be sustained by polysemy. A polysemic adjective like *civil* adjusts its meaning in accordance with the LCS of the noun whose intension it circumscribes. The specific meaning of the adjective is delimited by the lexical-conceptual relation between N and A. In (7) examples are given of four different types of meaning that Portuguese [NA] nominal compounds with the adjective *civil* can have:

(7) a. event: [[*guerra*]$_N$ [*civil*]$_A$]$_N$ 'civil war', [[*casamento*]$_N$ [*civil*]$_A$]$_N$ 'civil marriage'

b. state: [[*estado*]$_N$ [*civil*]$_A$]$_N$ 'marital status'

c. human institution: [[*polícia*]$_N$ [*civil*]$_A$]$_N$ 'civil police'

d. specialized professional domain: [[*engenharia*]$_N$ [*civil*]$_A$]$_N$ ‘civil Engineering’

The meaning of the adjective varies according to the meaning of the noun it modifies. The history of the entities highlights the semantic features focused in each case.

A *guerra civil* ‘civil war’ is a war between organized groups within a single nation. The adjective means, in this context, ‘intra-national’, by opposition to international. A *casamento civil* is a secular marriage, as opposed to a religious one. Since the Middle Ages, *civil* has been used in contrast to *ecclesiastic*. In both cases the adjective modifies an eventive noun; nevertheless, the meanings diplayed by the adjective are independent of the eventive class: they are correlated with the specific lexical meaning and profile of each noun.

The Brazilian *polícia civil* ‘civil police’ denotes the investigative state police forces. In this case *civil* is opposite to ‘military’ (cf. *polícia militar* ‘military police’).

The *estado civil* denotes the ‘marital status’, the legal standing of a person in regard to his/her marriage state. In Portuguese, the adjective *civil* covers, as a hyperonym, all the types of marital status: single, (un)married, divorced, widow(er).7

Civil Engineering is a hyponym of *engineering*. The adjective presents a technical meaning, referring not only, as in the past, to non-military and/or non-ecclesiastic engineering, but also to a wide variety of subdomains, including all the classes of construction engineering and construction materials. This hyponym term opposes *civil* to *electrical*, *biomedical*, *geological engineering*.

The semantic diversity and specialization of the adjective is correlated with the semantics of the noun it modifies. The history of the culture and of the society highlights the motivations of this variation.

3.3 $[NN]_N$: semantic frames and world knowledge

When two nouns form a compound, the meanings of both must articulate and construct a plausible meaning. As a large variety of semantic relationships is possible between the nouns,[10] a large frame of conceptualization and of reference is necessary to explain NN semantic profusion and diversity, namely when unexpected meanings emerge. But if a gap between the plausible meanings and the conventionalized one remains, then only a specialized source of information can provide the idiomatic meaning.

3.3.1 Proposals

A speaker uses compounds as memorized constructions whose holistic meaning is understandable, despite their degree of semantic idiomaticity. The speaker is able to use the word, even unknowing how the idiomatic meaning has been built. However, the speaker must understand the idiomatic meaning of the word if he intends to use it. When the comprehension of the meaning is not straightforward, which happens when the meaning of the whole is not literal and compositional regarding the meaning of the parts, what are the means applied to understand the word?

---

[10] Jackendoff (2010: 436-442) presents the following fourteen basic functions for English NN compounds:

(i) CLASSIFY (X, Y): 'N1 classifies N2' (*X-ray, Molotov cocktail*)
(ii) 'N2 of/by N1' (*food surplus, German grammar, used-car prices*)
(iii) BE (Y, X), 'Y is /also) an X' (*child prodigy, fisherman, woman doctor*)
(iv) SIMILAR (X, Y) 'an N2 similar to N1' (*tree diagram, zebrafish*)
(v) KIND (X,Y) 'an N2 of Kind/that is a kind of N1' (*bear club, ferryboat, girl child*)
(vi) BE (X, AT/IN/ON Y) 'N2 is located at/in/on N1' (*brain tumor, blood sugar, tree house*)
(vii) COMP (X,Y): 'N2 is composed of N1' (*brass instrument, ice sculpture, sheet metal*)
(viii) MADE (X, FROM Y) 'N2 made from N1' (*apple juice, coal-tar product*)
(ix) PART (X, Y) 'N2 is part of N1' (*computer screen, lunch meat, wheelchair*)
(x) CAUSE (X, Y) 'N2 that is caused by N1' (*sunburn, knife wound*)
(xi) MAKE (X, Y) 'N2 made by N1' (*honeybee, polio virus, songbird*)
(xii) X SERVES AS Y: 'N2 whose (proper) function is to function as an N1' (*endpoint, ferryboat, guard dog, guidebook*)
(xiii) HAVE (X, Y) 'N2 that as (an) N1' (*AIDS baby, gangster money*)
(xiv) PROTECT (α, Z, FROM X1) 'N2 protects N1 from something' (*lifeboat*) or N2 protects something from N1' (*sun hat, gas mask*).

First of all, the speaker tends to construe the meaning of each compound component. The semantics of each component is mentally constructed in accordance with the conceptual and the denotational representations associated with it.

According to its ontological nature, each N is characterized by a cluster of semantic or thematic roles that are connected with the LCS schemata associated with it and by the network of possible semantic relations the N can establish. A constellation of conceptual features (cf. e. g. Jackendoff 1990), like BE, DO, HAVE, SEEM, BEHAVE, CAUSE, which are the basic building blocks of LCS structures, profiles the semantics of a lexical item and, in consequence, the thematic relations it supplies (Dowty 1989, 1991; Jackendoff 2010).

In order to bring plausibility and transparency to the semantic relationship built by the compound, the speaker takes into account all the features and scenarios — the more and the less prototypical — associated with each word and its denotational frame: if necessary, even the possible semantic features of each word are mapped for this demanding computation task. World knowledge and/or referential coercion can also be used, as supported by a compatibility purpose. As Jackendoff (2002: 250) says, in order to determine the meaning of a newly encountered compound 'one uses the Head Principle, plus the repertoire of possible semantic relations, plus a dose of pragmatics, to put together a meaning that makes sense in context'.[11]

[11] According to Baroni et al. (2007), the noun combinatorial history influences the interpretation of a novel phrase involving that noun. That is, people use the distributional knowledge of how nouns have previously been combined to interpret a novel combination.

3.3.2 Semantic roles and the "maximal semantic compatibility principle"

Let us consider some Portuguese [NN] compounds in order to describe how coindexation underlies their semantic processing. We begin with four NN compounds lexicalized in the 1990s and later we analyse two novel/possible compounds.

Recent psycholinguistic research (Gagné and Shoben 1997, Gagné and Spalding 2006) emphasizes the relevance of compound interpretation based on thematic relations. Several sets of thematic roles have been proposed.[12] We adopt Rio-Torto and Ribeiro's (to appear) framework, with ten thematic relations that are suitable for Portuguese compounds with the structures $[N[PN]_N$, $[NN]_N$, $[NA]_N$. Here, four thematic relations are taken into account: container/recipient, goal, similarity, source**.** We claim that a dynamic framework of thematic roles is needed to explain their frequent intersections.

The meaning of the nominal modifier N2 and that of the nominal head N1 must coarticulate and converge into an output compatible with human categorization, with human world knowledge and with human experience. Coindexation involves and/or forces all the meanings — literal, denotational, figurative — so that a plausible meaning emerges.

(9) $[[bébé]^N$-$[proveta]^N]^N$ 'test-tube baby'

How should the typical characteristics associated to a baby with a test-tube, in $[[bébé]^N$-$[proveta]^N]^N$ 'test-tube baby', be reconciled? This container represents the

[12] Shamsfard and Mousavi (2008) work with seventeen roles: agent, experiencer, patient, theme, time, location, cause, source, destination, reason, topic, instrument, force, state, comparison, message, beneficiary. Jackendoff (2002: 250) announced a twenty roles repertoire for English NN compounds. Jackendoff (2010) presents a list of fourteen semantic roles.

'place'/'location' of fertilization. In fact, [[*bébé*]^{N}-[*proveta*]^{N}]^{N} 'test-tube baby' is a baby conceived by *in vitro* fertilization.

In comparison with professional, physical and mental properties, the containers do not represent salient features of a human being. But for specific types of human beings — for some fetuses —, a container such as a test-tube denotes something fundamental in their conception: an artificial womb where fertilization of the ovum took place and where the human embryo developed before being transferred to the mother's body. So, in this case the (artificial) place of fertilization replaces the biological container (the uterus) where conception typically occurs. World knowledge is here crucial for the recognition of the specific meaning of the word.

This example shows in a higher degree the denotation coercion imposed on the semantic relationship between N1 and N2, as [container] is not an expected prototypical feature of N1 when it denotes a baby. As with several containers, two semantic roles are also involved here, because the artificial fertilization occurs by means of a test-tube and in its interior.

(10) [[*criança*]^{N}-[*soldado*]^{N}]^{N} 'child soldier'

A child has no professional activity, and a soldier is a military professional. Nevertheless, the denomination *criança-soldado* 'child soldier' refers to children that act as and became combatants. It is the goal, the telic function that is focused. Once again, coindexation is forced to select a non typical and non-expected feature of N1.[13] Otherwise, this compound could be understood as denominating a child simulating a combatant, which is not in accordance with the extralinguistic reality.

13 We can admit that the coindexed features (AGENTIVE, CONSTITUTIVE, FORMAL, TELIC) belong to the qualia structure of each noun (Pustejovsky 1995). However, the selection of the specific information coindexed in each compound is governed by lexical-conceptual and referential guidelines of plausibility and correspondance with world knowledge.

(11) [[*homem*]^{N}-[*rã*]^{N}]^{N} ‘lit. man-frog’

In contrast, in *homem-rã*, prominence is given to the look, the similarity with a frog, in appearance and in functions.[14] A *homem-rã* is a diver, a person who explores underwater, especially equipped with breathing apparatus and weighted clothing.[15] As the similarity involves the appearance and the function, two thematic dimensions are also merged.

(12) [[*Retrato*]^{N}-[*robot*]^{N}]^{N} ‘photofit picture’

Retrato-robot ‘photofit picture’ refers to a product of a specialized method of combining photographs of facial features, hair, etc., into a composite picture of a face. This method is used by the police to trace suspects from witnesses' descriptions. Like a robot, that is, an electro-mechanical machine that is supposed to function as a human being, the photofit picture is obtained by computer means in order to create a schematic and artificial sketch of a real being. Regarding its semantic role, robot represents the robotic device by which the picture is drawn and/or the schematicity (robot-like) of the product. The specialized meaning is not available nor computable without a technical source of information.

For those with a global world knowledge about the denotation of these four [NN]N compounds, their interpretation offers no difficulties. It is not the case for a child, when not yet exposed to them, or for a speaker of Portuguese as a foreign

[14] As Jackendoff (2010: 429) emphasizes, one of the procedures involved in combinatorial semantics of NN compounds is profiling or topicalization, according to which a feature is picked out and is selected as the one to be referred to. We claim that the selected feature is present in the semantic frame of the units involved, and is the one that better fulfills the conditions of semantic compatibility between them.

[15] Contrarily to English, Portuguese compounds are typically left-headed: this explains the apparent shift of head between the Portuguese (*homem-rã*, lit. man-frog) and English compounds (*frog-man*) for the same concept.

language, especially if there are no directly corresponding expressions in their native language.

The computation task of understanding novel compounds, such as *bébé-brinquedo* 'lit. baby-toy', *mãe-polícia* 'lit. mother-police', highlights that (i) several possible semantic dimensions can be gathered for the understanding of the whole and that (ii) the conventionalized meaning of each word is not necessarily reached without the help of the co(n)text and/or world knowledge. A *bebé-brinquedo* can be a *bebé* 'baby' used as a 'toy' by their brothers, or a *bebé* that evokes a toy, because of their way of moving or playing or even a toy, like a doll, imitating a real 'baby'. A *mãe-polícia* can be a *mãe* 'mother' whose professional activity is a *policewoman*, but also a *mother* that controls and/or excessively protects her children. In both cases more than one thematic role can be involved and the precise meaning of the compound results from the interaction between linguistic information and devices and extralinguistic sources of knowledge.

When a compound has several readings,[16] multiple semantic-roles must be used to describe it. Thematic relations, since conceived dynamically and interacting with world knowledge, can provide the compound with semantic plausability.

3.4 Conclusions on compounding

The semantic structure of each compound reflects the 'maximal semantic frame' associated with each of the constituents, as well as the plausible semantic and

[16] Cf. *boxcar* 'car that carries boxes, that resembles a box, that serves as a box' (Jackendoff 2010: 428).

grammatical relations relying on them.[17] Coindexation, in accordance with a semantic plausibility principle, assures the maximal compatibility between the meanings involved.

The specific meaning of each compound is due to the semantic structure of both constituents, in articulation (i) with semantic/conceptual templates governing the relation between compounding terms and (ii) with referential, pragmatic or figurative constraints. The floating adjustments in the meaning of compounds are mainly governed by referential and/or pragmatic motivations. Figurative mechanisms provide semantic coherence when denotational or objective tools are overlooked.

From a production point of view, a compound is a construction whose meaning is anchored, to a variable degree, to the meaning of its constituents. But because of referential and/or pragmatic reasons, idiomaticity and opacity affect the compound, and for this reason the interpretation can be only weakly or scarcely compositional.

Due to the LCS frames associated to each compound member, and due to the semantic features that result from their combination and by reference and plausibility needs, the range of semantic possibilities of the final meaning, though not unlimited, is quite broad.

A theory of lexical (de)composition must incorporate a textured set of dimensions and procedures that cut across lexico-conceptual representations,

[17] Jackendoff says (2009:250) that in compounding «the grammatical principled involved is simply one of concatenating two nouns into a bigger noun, and the semantic relation between them is determined by a combination of pragmatics and memorization». Despite the apparent simplicity of grammatical relations of coordination, subordination or modification between the nouns of a compound, pragmatic, referential and/or conceptual informations associated to each noun play a major role in the construction of the conventionalized meaning of the whole.

coindexation, referential coercion, semantic shifts and figurative devices of meaning production.

4 CONCLUSIONS

Are there radical differences between the semantic processing of derivations and compounds?

i) Derived words, as well as compounds, can present compositional and idiomatic meanings. In both cases the whole can display a compositional meaning or an idiomatic one, not computable from the meaning of the parts.

ii) In both cases semantic coindexation is responsible for the construction of the meaning, supported by the articulation between the semantics of the units and their maximal semantic frame.

iii) In both cases the meaning of the whole respects semantic coindexation guidelines of maximal/minimal compatibility between the features of the lexical units involved and, respectively, the most prototypical/the less representative meanings and nouns of the paradigm. In both cases a straightforward compatibility leads to a transparent noun; the more flexible or non linear the meaning construction is, the higher the idiomaticity.

iv) Nevertheless, the set of LCS features associated with a lexeme, and, *a fortiori*, with a relation between two lexemes, is potentially wider than the set of features associated with a suffix. This difference opens up the universe of possible semantic and denotational codifications performed by a compound. So, due to the fact that compound constituents represent two LCS universes that articulate and enrich reciprocally, the meaning structure in compounding tends to be freer than in

derivation. However, the relation between the nouns is not totally free. By default, the meaning of a derived word is less unpredictable than the meaning of a compound.[18]

[18] We could equate this observation with Jackendoff (2010: 422-423) proposal that compounding is a relic of protolanguage, that is, an architecture anterior to language that would contain semantics and phonology, but no syntax. According to Jackendoff, compounding (but not derivation) displays some properties of a 'protolinguistic' fossil, namely a rudimentary grammatical structure that does not shape semantic interpretation.However, we claim that the alleged freedom of compounds meaning is indeed constrained by the semantics of the units involved and by pragmatic and/or referential conditions of compatibility between them. What we question is: if semantics, as Jackendoff proposes, has a generative character, and we believe it has, according to empirical data analysed in our works on word formation, would that generative character be present in the semantics of protolanguage? Would protolanguage semantics be so complex and structured as it is language semantics?