WORD-INTERNAL SYNTACTIC COMPLEXITY

COMPLEXIDADE SINTÁTICA NO INTERIOR DAS PALAVRAS

Isabella Lopes PEDERNEIRA
Universidade Federal do Rio de Janeiro (UFRJ)

Miriam LEMLE
Universidade Federal do Rio de Janeiro (UFRJ)

ABSTRACT

The purpose of this study is to examine the successive steps of complex word formation in Portuguese, with special attention to how big can words become and how compositional and idiomatic readings can alternate. The question of where idiomatic readings can occur and alternate with compositional readings in big words is crucial for the development of the Theory of Grammar. These grammatical aspects of the word-internal syntax-semantics interface will be meticulously analyzed, with special accuracy in the analysis of semantic alternations in verbs according to their syntactic contexts. Such empirical data will be decisive for theoretical options between alternative theoretical hypotheses concerning non-lexicalist models.

RESUMO

O objetivo deste estudo é analisar as etapas sucessivas na formação de palavras complexas em português. Pretendemos aprofundar a questão da alternância entre leituras composicionais e leituras idiomáticas dentro das palavras complexas. A posição de incidência e a alternância de leituras idiomáticas em palavras longas é crucial para a formação da teoria gramatical. Aspectos gramaticais da interface sintaxe-semântica no interior das palavras serão analisados meticulosamente com especial atenção para a análise de alternâncias semânticas em verbos segundo seus contextos sintáticos. Dados empíricos dessa natureza serão decisivos para as opções teóricas entre versões alternativas de modelos não lexicalistas.

KEYWORDS

word-internal constituent structure; word-internal syntactic complexity; word-internal compositional readings; word-internal idiomatization.

PALAVRAS-CHAVE

Estrutura de constituintes em palavras; complexidade sintática em palavras; leituras composicionais em palavras; idiomatizações em palavras.

Introduction

The purpose of this article is to highlight the richness of word-internal syntactic formation, with a special interest on its property of repeated application of rules of word-internal categorization. Morphology in Portuguese is very transparent with respect to word-class marking. For example, if a noun ends in –*eza*, the speaker knows that he has to grasp an adjective inside it. If an adjective ends in –*ento*, the speaker recognizes a noun inside it. The morphological identification of a predicate pops up, independently of meaning, from the determination of the tense/agreement feature bundle in the verb. Word-formation may be reapplied repeatedly by merging morphological word-internal pieces that produce multi-layered words that have to be read stepwise by small semantically regular compositional calculi. For Portuguese speakers it is intuitively obvious that syntax goes all the way down into words.

We will present Brazilian Portuguese data that illustrate that there is abundant Syntax within words, and raise a problem with respect to the ciclicity of semantic readings. The demonstration that words are syntactically analyzable will be organized from simple word patterns to configurations of complex verbs. The merit of our observations of the syntactic-semantic relationships inside words is that they will favor the constructionist hypotheses with respect to projectionist hypotheses

and also highlight how compositional readings and idiomatic ones can both exist for one and the same complex word. Within constructionist alternatives, the Exoskeletal theory (Borer 2005a) will be shown to work better than Distributed Morphology (Marantz 1997).

1. Theoretical Dissidences

The empirical data we are going to present and discuss are relevant for theory formulation in what regards the lexicon-syntax interface. The basic dissidence in this area is the lexicalist X constructionist one. According to Chomsky (1970) the consideration of semantic idiosyncrasies in deverbal nominalizations leads to the radical separation between verbs and their derived nouns. This separation results in a model of Grammar that came to be called Lexicalism (HALLE, 1973; ARONOFF, 1976), in which the Lexicon and the Syntax are distinct modules.

The Distributed Morphology (DM) proposal in Halle & Marantz (1993) claims that Syntactic Derivation reaches "all the way down" into words. Marantz (1997) considers morphology as word-internal Syntax, especially because of the wide occurrence of multiple layered words with compositional meanings. It is observed that, in these words, the idiosyncratic part of the meaning normally occurs in the first categorization. Marantz proposes this empirical fact as a generalization, and therefore states that the place of idiomatic formation is the first categorization slot of complex words, as in *globalization*: $[[[[glob]_n al]_a iza]_v tion]_n$, where the conventionalized meaning is at the nominal most internal component, *globe*.

A second opponent of Lexicalism is Hagit Borer's Exoskeletal (X-skeletal) model (Hagit 2005a), which agrees with DM in the *syntax all the way down* observation, but opposes it regarding the place where semantic idiosyncrasy may incide. As a first empirical argument, Borer offers the word *reactionary*, which has the property of being an adjective

with a meaning not compositionally derived from the meaning of its internal noun *reaction*. Besides this, the word *reaction* on its turn does not receive a compositional meaning with respect to *react*. Borer's conclusion from facts like these leads her to build a theory in which syntax and semantics are more radically dissociated. In her theory, the word-internal syntax-semantics interface is relevant, with encyclopedic search being allowed to apply at any layer, and the meaning of roots devoid.

The Portuguese language used in Brazil is an especially good source of data for deciding the best theory, due to its possibility of deep syntax inside words, and wide variation of positions in which word-internal idiosyncratic semantics can fall, as, for example in *construtivismo constructivism, construcionismo constructionism, cotovelada kick at or with elbow, chicotada lashing, colherada spoonful, reacionário reactionary, governanta housekeeper*, etc.

2. Word Categorization and Recategorization

We are adopting the theoretical assumption that roots on their own are uncategorized and possibly meaningless listed units. It is in the Syntactic derivation that category functional morphemes merge to roots to form words, as proposed in Marantz (1997). Classifier morphemes can be nominalizers, verbalizers and adjectivizers. The shortest nouns consist of roots followed by the thematic vowels *–a* (*casa*) house, *-e* (*nome*) name, *-o* (*ninho*) nest, or by no vowel (*mar sea, sol sun, cruz cross*); the shortest adjectives, of roots followed by the same three vowels or no vowel (*bonito handsome, bonita beautiful, forte strong, central central, regular regular, feroz wild*); the shortest verbs, of roots plus the thematic vowels *-a*, *-e* and *-i* (*cantar sing, bater beat, dormer sleep*). Semantic arbitrarity is the general interface principle for one layered words such as these.

Each one of these word-classes can be a starting point to recategorization by category-forming morphemes of every other class, except suffixal verb-to-verb direct derivation.

In the examples of word-to-word derivations given below, the first one shows compositional meaning and, in the case there is an additional idiossycratic meaning, it is described after the compositional readings and written in *italic*.

1)   Noun to noun:
      Compositional reading (CR):
      brinco (*earring*)>brincão (*big earring*)
      pena (*feather*)>penugem (*fluff*)

      Idiomatic reading (IR):
      bola (*ball*)>bolinha (*small ball or synthetic drug*)
      carta (*letter*)>cartaz (*poster*)
      carta (*letter*)>cartilha (*primer*)

2)   Noun to verb:
      CR: telefone (*telephone*)>telefonar (*to make a phone call*)

      IR: coroa (*crown*)>coroar (*to crown*)
      cola (*glue*)>colar (*to glue or to copy in an exam*)
      grampo (*clip*)>grampear (*to clip or tapping the telephone*)

3)   Noun to adjective
      CR: carinho (*affection*)>carinhoso (*affective*)
      Chile (*Chile*)>chileno (*Chilean*)

      IR: cabelo (*hair*)>cabeludo (*hairy; in the context of joke, dirty*)

4)   Verb to noun:
      CR: tratar (*to treat*)>tratamento (*treatment*)
      trair (*to betray*)>traição (*betrayal*)
      montar (*to ride or to assemble*)>montagem (*mounting or assembly*)

IR: partir (*to depart or to split*)>partida (*departure or match*)
fritar (*to fry*)> fritada (*fry*)
criar (*to create or to raise*)> criado (*servant*)
protestar (*to protest*)> protestante (*protestant*)

5) Verb to verb:
CR: contar (*to count or to count on or to tell*)>recontar (*to count again or to tell again*)

IR: contar (*to count or to count on or to tell*)>descontar (to *discount*)
saltar (*to jump*)>ressaltar (*to emphasize*)
falecer (*to die*)>desfalecer (*to faint*)
pedir (*to ask*)>despedir (*to dismiss*)

6) Verb to adjective:
CR: dobrar (*to fold* or *to double*)>dobrado (*folded or doubled*)
dividir (*to divide*)>divisor (*divider*)
falar (*to speak*)>falante (*speaker*)
repelir (*to repel*)>repelente (*repellent*)

IR: fritar (*to fry*)>frito (*fried* or *in a big trouble*)
pender (*to recline*)>pendente (*pending*)
competir (*to compete*)> competente (*competent*)

7) Adjective to noun:
CR: duro (*hard*)>dureza (*hardness*)
idiota (*idiot*)>idiotice (*stupidity*)

IR: mesmo (*same*) > mesmice (*monotony*)
excelente (*excellent*)>excelência (*excellency* or *Excellence*)
salgado (*salty*) >salgadinho (*hors d'oeuvre*)
fresco (*fresh*) >frescão (*air conditioned bus*)
santo (*saint*)>Santíssimo (*God*)

8)   Adjective to verb:
      CR: molenga (*easygoing*)>molengar (*act in an easygoing manner*)
      forte (*strong*)>forçar (*to force*)

      IR: amarelo (*yellow*)>amarelar (*to act cowardly*)
      sujo (*dirty*)>sujar (*to dirty* or *to go wrong said of an arrangement*)

9)   Adjective to adjective:
      CR: duro (*hard*)>duríssimo (*very hard*)

      IR: bobo (*silly*)>boboca (*naive*)
      magro (*thin*)>magrela (*thin, up to weakness*)

We would like the reader to realize that, at the second categorization layer, derived words can have either regular compositional reading or idiomatic semantic readings. For example, in the Noun to Noun (1) derivations, *brincão* means *big earring*, *penugem* means *small feather*, but *bolinha* may mean both *small ball* or *synthetic drug* and *cartaz* has no meaning related to *carta* (*letter*) but means *poster*. In the examples of Noun to Verb (2) derivations, *telefonar* means *to use the telephone for its normal purpose of talking at a distance*. *Coroar* means *to crown*. *Colar* may mean *to glue* because *cola* means *glue* or also it can mean *to cheat at school by copying from someone else's exam*. *Grampo* means *clip*, but *grampear* can mean both *hold by clipping* if the direct object refers to a clippable thing or *telephone tapping* if the verb's direct object is the word *telephone*.

From (3) to (9) the reader should pursue the same kind of distinction between compositional and idiosyncratic readings in category-changing syntactic derivations.

Recategorization consists of the addition of a new category-marking morpheme to a word, possibly a suffix or some other formal procedure

of category marking. A regularly formed new meaning corresponds to each category change. However, in addition, an idiosyncratic new meaning might be formed. At this point, we are on the verge of new understandings of the syntax-semantics interface. As we started to point out in paragraphs 2 and 3 of Theoretical dissidences, the DM and X-skeletal models diverge on their hypotheses: Marantz (1997) states that semantic idiosyncrasies occur always on the first categorization syntactic level of complex words. Borer (2009) furnishes examples of semantic idiosyncrasies falling at late structural level of word formation. Since these hypotheses are totally incompatible, we have made a data collection of complex words to resolve the conflict:

10)   ferruginoso $[[[\sqrt{\text{ferr-u}}]_n \text{gin}]_n \text{oso}]_a$
11)   carroceria $[[[\sqrt{\text{carr-Ø}}]_n \text{oç}]_n \text{eria}]_n$
12)   americanizar $[[[\sqrt{\text{americ-a}}]_n \text{n}]_a \text{izar}]_v$
13)   globalização $[[[[\sqrt{\text{glob-Ø}}]_n \text{al}]_a \text{iza}]_v \text{ção}]_n$
14)   amabilidade $[[[\sqrt{\text{am-a}}]_v \text{bil}]_a \text{idade}]_n$
15)   importância $[[[\text{im}[\text{port-a}]_v]_v \text{nt}]_a \text{ia}]_n$
16)   governador $[[[\text{govern-a}]_v \text{d}]_{pp} \text{or}]_n$
17)   governanta $[[\text{govern-a}]_v \text{nta}]_n$
18)   canelada $[[\text{canel-Ø}]_n \text{ada}]_n$
19)   cotovelada $[[\text{cotovel-Ø}]_n \text{ada}]_n$
20)   colherada $[[\text{colher}]_n \text{ada}]_n$

Considering the examples above, some speakers may not be acquainted with the idiomatic meaning of *rust* in the word *ferrugem* in (10), or the idiomatic meaning *gentle* instead of *lovable* for the adjective *amável* in (14). It is interesting to note that after the idiomatization point at the deverbal adjective *amável*, a new layer may be added forming the noun *amabilidade*, and this layer has compositional reading relative to the *gentle* meaning of *amável*.

These findings of idiomatizations internal to complex words are crucial cases for the decision about the place of the Encyclopedia in the model of grammar. The examples (15)-(20) favor this idea: we have ambiguous words that may get both a purely compositional meaning and an idiomatic reading. The noun *importância* in (15) contains the adjective *importante*, which corresponds to the verb *importar*, which contains the verb *portar*. The noun *importância* may mean both, compositionally, importance and, idiomatically, 'amount of money'. *Governador* in (16) means *governor*, and refers to a person that governs a state; *governanta* in (17) means *housekeeper*, referring to a person who governs a home. In (18), (19) and (20), *canela* means shin, *cotovelo*, elbow and *colher*, spoon. *Canelada* means a blow received in the shin, *cotovelada*, a poke made with one's elbow and *colherada*, a spoon sized measure. Therefore a speaker of Portuguese has to know not only the arbitrary meanings of *canela*, *cotovelo* and *colher* but also different conventional relationships between each word ending in –*ada* and their respective readings. The set of examples (15)-(20) favors Borer's hypothesis about *Encyclopedic search* being allowed to be activated in the middle of the computation of a phonological word and disfavors Marantz's in which the *Encyclopedia* is believed to be activated only at the very first categorial marking in a word's computation.

The verb class is especially interesting, since semantic variation is wide in many verbs, and therefore we need to clarify its factors and limits.

The cases of verb polysemy that we take in consideration in this investigation are only the structure dependent ones, and exclude metaphors and idiomatic expressions.

We will present three examples for this syntax-semantics interface analyzing the different senses of the verbs *colar* (*to glue/to copy in an exam*), *montar* (*to assemble/to ride*) and *render* (*to yield/to produce/to surrender*).

(i)   The verb *colar* may appear in the context [Voice [V PP]] in which it means *to attach by gluing* as in:

Antônio colou o selo na carta
Antônio glued the stamp on the letter

In this context the PP complement has the syntactic structure seen in the representation: $[[o\ selo]_{DP_i}\ [n\ [a\ carta]_{DP_j}]_p]^1_{PP}$. Inside the PP, the reading of DP_i, in the terminology of theta-roles, is theme and the reading of DP_j is target.

(ii)   *Colar* may appear also in an unaccusative context [DP V] as in:

Este selo cola bem
This stamp sticks well

In this context, the subject DP is read *a thing that gets into the final state of glued,* because the verb is denominal, derived from *cola* (glue).

There is a second intransitive context with the form [Voice V] in which the verb means 'cheating at school by copying' where the DP subject is an agent as in:

Antônio colou na prova
Antônio cheated at the exam by copying

(iii)   The meaning of 'cheating'may emerge also from a transitive context [Voice [V DP]]:

Antônio colou a segunda questão
Antônio cheated by copying the second question

---

[1] In traditional grammar terminology, DPi is called the direct object and DPj is called a place adjunct

(iv)   In this extended example, the DP object may optionally be
        followed by an adjunct PP, to be read as source:

Antônio colou a resposta da prova do amigo
Antônio copied the answer from the friend's exam

Note that the resulting structure [[Voice [V DP]] PP] is the one that
represents the extended sentence (iv).

In this analysis, we have discovered that there are on the whole four
structures for the verb with phonological form *colar*. However, these
four structures are distributed in two pairs of two structures for each
verb (the *gluing* and the *copying*). For the verb *colar* with the meaning *glue*
the essential part of the meaning is in the unergative structure [DP V],
este selo cola bem, where the DP is the entity which has the power of *gluing*.
The agent of the event merges at the voice component [Voice [V PP]],
as in *Antônio colou o selo na carta*.

For the idiomatic meaning 'copy in school work', the anti-ethical
agency of the cheating action is the essential part of the meaning in
the structure [Voice V], *Antônio colou na prova (Antonio cheated in the exam)*.
In the transitive version, [Voice [V DP]], *Antônio colou a segunda questão
(Antonio copied the second question)* the direct object describes the copied
material.

There are actually three structures. Transitive, unaccusative and
unergative. Whilst transitives are possible with both the gluing and the
school-cheating idiomatic readings, unaccusatives are only possible with
cheat idiomatic readings, and unergatives, with conventional *glue* readings.

In these examples each sub-case of the two meanings emerges
from different syntactic structures. The essential structure of the gluing
meanings is the intransitive in which the subject is read as the thing that
gets glued. The essential structure for the cheating/copying meaning is

the voice component. So, we see that the polysemy of the verb *colar* (*to glue* or *to copy*) depends on the syntactic structures [DP v] *versus* [Voice v], respectively.

Now, let us see the syntax-semantics interface of *montar* (assemble/ride):

(v)   In the transitive construction [Voice V DP] the verb means to assemble, as in:

Meus netos montaram um quebra-cabeça enorme
My grand-children assembled a huge puzzle

(vi)   In the transitive with prepositional complement construction [Voice V PP] the verb means to ride.

Antonio montou no cavalo
Antonio rode the horse

In the transitive structure the verb *montar* assumes a notion of creation, and in the prepositional phrase structure it realizes a relation of localization in which the sentential subject is understood as the located entity and the complement of the preposition *em* (n-) is read with the role of location.

As a third example, let us look at the verb *render:*

(vii)   In the intransitive construction [DP V] the verb captures the notion of growth of the thing referred to by the sentential subject.

Esta poupança rendeu
This savings account yielded

(viii) A transitive construction [DP V DP] informs, in the object
DP, the measure of yield of the subject referent:

Esta quantidade de massa rende dois bolos
This amount of dough yields two cakes

In (vii) the verb describes a process undergone by the sentence
subject, and in (viii) it describes a result, the final state attained by the
subject DP.

(ix)   In a voice transitive construction [DP$_i$[Voice [V DP$_j$]]] the
verb *render* means "surrender".

O assaltante rendeu a moça
The burglar made-surrender the girl

The polissemy of this verb is quite intriguing, especially with regards
to the leap between (vii)-(viii) and (ix). How can we explain the semantic
leap between the readings of structures (viii) and (ix) without appealing
to homophony? In (ix), if we select human-referring nouns for DP$_i$ and
DP$_j$, there will be only one reading possibility for the two DPs: you have
to read DP$_i$ as an agent relative to DP$_j$, and read DP$_j$ as a theme. In this
case, the *yielding* sense present in (viii) can be satisfactorily implemented
by imagining for DP$_j$ a use condition including a consciously surrended
entity.

Summarizing: for the *colar*, *montar* and *render* syntactic contexts, each
context licenses different meanings. The observation that different
meanings depend on different syntactic structures is theoretically

important: if the meaning was derived from the root it would not vary according to syntactic context, since the root itself would impose its meaning. And if a root's meaning was an essential feature for the reading, then we would have to postulate more than one lexical entry for homophonous roots. This conclusion favors Borer's hypothesis about the semantic poverty of roots.

We are going to analyze complex structures in other Portuguese verbs in order to contribute to the understanding of the mechanics of semantic formation in the derivation of verbs from nouns, verbs from adjectives, verbs from prefix+noun, prefix+adjective, prefix+verb.

3. Complex Strucure in Portuguese Verbs

Verbs are especially interesting because they can merge in many different syntactic contexts and therefore they present many hypothetical syntactic layers to check idiomatization possibilities.

The minimal verb formation possibilities are:

(21) Direct categorization of bare roots: √cas+ar *marry*, √ca+ir *fall*, √bat+er *strike*, √diz+er *say*, √v+er *see*, √forn+ecer *furnish*, √sintet+izar *synthesize*, √corr+er *run*, √part+ir *divide/depart*.
Note:The meaning of the structure resulting from the first categorization of a root is necessarily arbitrary.

(22) Recategorizing nouns:
Compositional readings (CR): $[[\sqrt{martel}+\emptyset]_n\ ar]_v$ *hammer$_n$, to hammer$_v$.*
The same structural description is found in the example that follow: pentear *comb, to comb*; escovar *brush, to brush*; olhar *eye, to look*; ladrilhar *tile, to put tiles*; arquivar *file, to file*; manejar *hand, to manipulate*, farejar *flair, to smell*; perfumar *parfum, to parfum*.

Note: The semantic relation between the thing named by the verb-internal noun and its use referred by the verb is regular. The regularity consists in the fact that if you know what the thing mentioned inside the verb is used for you know automatically that the meaning of the verb is 'using the thing appropriately'.

Idiomatic readings (IR):[[√alfinet+Ø]$_n$ar]$_v$ *pin$_n$, to offend subtly$_v$.* Incensar *incense, to incense*; grampear *clip, to tap*; lixar *sandpaper, to feel uninterested*; bolar *ball, to plan*; cornear *horn, betray the sexual partner*; pipocar *popcorn, many people or things appearing suddenly and briefly.*

Note: From this point on, for all verbs placed in the idiomatic class there is also a compositional meaning obtained by the appropriate calculus, which has the form [act with/on/like DP]. In the lists that follow, only the internal noun and the idiomatic reading will be mentioned:

(23) Recategorizing adjectives:
CR: [[√limp+Ø]$_a$ar]$_v$ *clean, to clean.*
Santificar *saint, to sanctify*, tranquilizar *quiet, to quiet down*, clarear, *clear, to clear up*; helenizar *helenic, to helenize.*

IR: [[√suj+Ø$_a$] ar]$_v$ dirty, *go wrong.*
Amarelar, *yellow; to give up.*

(24) Verbalizing the merge [prefix+verb]:
CR: [des+[√cas+ar]$_v$]$_v$, *to get married, to break a marriage,* condizer *to say, to make compatible*, refornecer *to furnish, to furnish again*, ressintetizar *to synthetize, to make a second synthetization.*

IR: [re[√bat+er]$_v$]$_v$ *to beat, to reply*;
combater, *to beat*, to *fight against*; rever *to see, to reconsider*; repartir, *to leave, to divide for distributing pieces*; ocorrer, *to run, to occur*; desandar, *to walk*, in cooking jargon, said of *any mistaken blend of ingredients*; discorrer, *to run, to talk at length about a subject matter.*

(25)   Verbalizing the merge [prefix+noun]:
CR: [[en+[lat Ø]$_n$]ar]$_v$ *can, to can*;
amassar *mass, to squeeze a mass*; aninhar *nest, to put in nest*; espernear *legs, to shake legs*; desossar *bone, to debone*; engaiolar *cage, to put in cage*; incorporar *body, to incorporate*; abotoar *button, to button up*; despontar *point, to appear*; apontar *point, to point*; soterrar *earth, to bury*; abraçar *arm, to embrace*; ajeitar *manner, to put in order*; apavorar *scare, to scare*; aterrorizar *terror, to terrify*; amedrontar *fear, to frighten*; apaixonar *passion, to become passionate*; engatinhar *cat, to crawl*; empedrar *stone, to become stony.*

IR: en+[gat]$_n$ar]$_v$ *cat, to clamp*
engavetar *drawer, paralize a bureaucratic process*; desvendar *veil, to unveil*; despencar *bunch, to fall down*; acarear *face, to confront witnesses*, encorpar *body, to thicken.*

(26)   verbalizing the merge [prefix+adjective]:
CR: [[em+√grand+Ø]$_a$+ecer]$_v$ *big, to turn big*;
engordar *fat, to fatten*; emagrecer *thin, to become thin*; apodrecer *rotten, to rotten*; arredondar *round, to turn round*, amolecer *soft, to turn soft*; amaciar *fluffy, to turn fluffy*; acovardar *coward, to become coward.*

IR: [en [√gross+Ø]$_a$ar]$_v$ *thick, to behave aggressively*;
amaciar (o motor) *soft, drive softly during the first ten kilometers of a new car*; amortecer *dead, minimize the hardness of a fall*; emburrar

stupid, look bad-tempered; enfadar *destiny, to bore*; encarecer *dear, to beg.*

At this point the reader should have enough data to realize the difference between the data in the previous section (from i to ix) and in this section (from 21 to 26).

In the previous section we excluded metaphors and idioms, and therefore it is by pure regular calculi that the Portuguese speakers construct the meaning of each verb.

In this section - from 21 to 26 - semantic variation going from compositionality to idiosyncrasy - is affected by a variety of use conditions.

Some of the verbs generated in (22)-(26) can be the input for subsequent verb formations by addition of a prefix to the previously formed verb:

(22a) CR: [RE [[√gramp+Ø]$_n$ ear]$_v$]$_v$, *clip, to clip, to clip again* DES+pentear (*comb, to comb, to get hair in disorder*), RE+escovar (*brush, to brush, to brush again*), RE+arquivar (*file, to file, to file again*) / DES+arquivar (*file, to file, to take out of file*)

IR: [RE+[[√bol+Ø]$_n$ ar]$_v$]$_v$ (*ball, to invent, to shake the hips*); Desnortear (*north, to give a direction, to get lost*);

(23a) CR: [DES+[√elen+izar]$_v$]$_v$ (*to hellenize, to deshellenize*) Relimpar (*to clean, to clean once more*)

IR: [[DE+[√negr+Ø]$_a$]$_{DE}$ ^{2}ir]$_v$ (*nigger, to denigrate*)

---

^{2} DE-, prefix, means from.
Nigger, from nigger, evaluate from notion of darkness.

(24a) CR: $[RE+[DES+[\sqrt{cobr}+ir]_v]_v]_v$ (*to cover, to discover, to rediscover*)
Coocorrer (*to run, to occur, to cooccur*), decompor (*to put, to compose, to decompose*), reatirar (*to extract, to throw, to throw again*)

IR: $[COM+[PRO+[\sqrt{met}+er]_v]_v]_v$ (*to put inside, to promise, jeopardize someone's reputation*),
Desenvolver (*to turn back, to involve, to develop*)

(25a) CR: $[DES+[A+[\sqrt{mass}+\emptyset]_n+ar]_v]_v$ (*mass, to squeeze, to unsqueeze*)
Desempedrar (*stone, to become stony, to undo the stony state*), desabotoar (*button, to button, to unbutton*)

IR: $[DES+[EN+[[\sqrt{rol}+\emptyset]_n+ar]_v]_v]_v$ (*roll, to roll, to pack, to persuade*)

(26a) CR: $[DES+[A+[\sqrt{rredond}+\emptyset]_a+ar]_v]_v$ (*round, to round, to undo a rounding*)
Reamolecer (*soft, to soften, to become soft again*), desamaciar (*smooth, to soften, to undo a smoothing*)

IR: $[RE+[EX+[\sqrt{fri}+\emptyset]_a+ar]_v]_v$ (*cold, to become cold, to get a cold*)

Once more the structures above can be input to one or more than one merge of prefix:

(27) Prefix+prefix+prefix+verb: $[RE+[DES+[EN+[\sqrt{gavet}+\emptyset]_n+ar]_v]_v]_v$ (*drawer, to put in drawer, undo put in drawer, once more undo put in drawer*)
Desimpermeabilizar (*permeable, impermeable, impermeabilize, undo the impermeabilization*).

(28) Prefix+prefix+prefix+verb:
[RE+[DES+[[IN+[√compat+ibil]$_a$]+izar]$_v$]$_v$]$_v$ (*compatible*,
*incompatible, to cause incompatibility, to undo causing incompatibility, to
cause return to a state in which incompatibility had not happened*)

What we see from (22) to (28) is that there is syntactic complexity
in word formation both in suffixation processes and in prefixation
processes. This property brings together sentential syntax and word-
internal syntax in quite a high proportion.

Another mechanism to form new verbs has intermediate steps of one
or more than one non-verb components between two verb components,
as in (29)-(32), where compositional readings may be calculated stepwise
up to the highest layer:

(29) contabilizar, v, *to put something in accountable state*
[[[[cont]n a]v bil]a izar]v
conta, n, *arithmetic operation*→contar, v *count* →contável, a *that
can be counted*→contabilizar, v *to put something in accountable state*

(30) disponibilizar, v, *make available*
[[[dis[pon] v]v ibil]a izar]v
por, v *put*→ dispor, v *offer for use*→ disponível, a *available*
→disponibilizar, v, *make available*

(31) avermelhar, v, *to redden*
[[a[vermelh]a]$_{prep}$ar]$_v$
vermelho, a. *red*→ a+vermelho, *to red*→ avermelhar, v, *to redden*

(32) ver, v, see; visto, ver, past participle; [[a[[vis]$_v$ t]a] PP ar]v
avistar, v, *to see from a distance*

A non-compositional meaning formation may happen preceding the highest layer's verb, as in (33) and (34):

(33)   cumprimentar, v, *greet*
    [[[cumpri]v ment]n ar]v
    cumprir,v, *accomplish an obligation* → cumprimento, n, *greeting (idiomatic meaning)* → cumprimentar, v, *greet*

(34)   regulamentar, v, *establish rules*
    [[[[[regula]n]v] ment]n ar]v
    (regula, n, *rule* LATIN) → regular, v, *create rules*→regulamento, n, *statute*→regulamentar, v, *enact statute*. Non-compositional meaning formation in the verb *regulamentar* happens at the step where the noun *regulamento* is formed from the verb *regular*.

Still another derivational pattern goes from a verb to another verb passing through the past participle of the first verb as in (35)-(41):

(35)   liberar, v, *to free* → liberto, a, *free* → libertar, v, *to free*
(36)   dizer, v, *to say* → dito, a, *said* → ditar, v, *to dictate*
(37)   ver, v, *to see* → visto, a, *seen* → avistar, v, *to see from a distance*
(38)   expelir, v, *to expel* → expulso, a, *expelled*→ expulsar, v, *to expel*
(39)   conceber, v, *to conceive* → conceito, n, *concept* → conceituar, v, *to conceptualize*
(40)   rever, v, *to see again* → revisto, a, *revised* → revistar v, *to inspect*
(41)   frigir, v, *to fry* → frito, a, *fried* → fritar, v, *to fry*

The derivations of (35)-(41) have been experimentally tested in (Pederneira 2010; Lemle & Pederneira 2012). The purpose of the test design was to compare the priming effects of three types of morphological relations between words, one of which was precisely of the type rever – revistar (to see again – to inspect), in which the second verb is derived

from a morphologically exceptional past participle. The result of the test was that the past participle-to-verb family of morphological relations had the weakest of the priming effects, approaching the control group. This derivation seems to work as if the step from a past participle to a new verb was as obligatorily idiosyncrasic as the first step of any word formation.

Conclusion

The numerous cases of compositional readings in long word derivations are a clear proof that "syntax goes all the way down". Syntax going all the way down is not the favored expectation for a theory in which lexical complex structure formation should be dissociated from syntactic computation. The fact that compositional readings exist in all derivational stages towards the formation of big words is a proof that complex word formation happens in the syntax. This is equivalent to saying that Constructionist theories hold the best prediction in comparison to Projectionist theories.

Among Constructionist theories alternatives, a new question raises: what are the possible incidence points for idiomatic readings inside complex words? The possible answers to this question may be either that there is only one possible point for idiomatic reading in complex word formation (Marantz 1997) or that all recategorization points may be loci for idiomatic reading formation (Borer 2004, 2005a, 2005b and 2009). We discovered and analyzed cases of idiomatic reading formation at midway in the verb derivation. This kind of phenomenon is not predicted by Marantz's expectations: the first layer of merge of root with categorizer functional morpheme as the only possible one for word-internal idiomatization; subsequent phases with compositional meanings. However, Borer's conception of Encyclopedic search predicts our findings: we have shown that Portuguese morphologically complex nouns, adjectives and verbs may receive an idiomatization at a morphological later merge than that of their first categorizer morpheme.

References

BORER, Hagit. **The Grammar Machine**. In: Artemis Alexiadou; Elena Anagnostopoulou & Martin Everaert (eds.). The Unaccusativity Puzzle. Oxford: Oxford University Press: 288-331. 2004.

_____. *In* **Name Only.** Structuring Sense, Vol. I. Oxford: Oxford University Press. 2005a.

_____. **The Normal Course of Events**. Structuring Sense, Vol. II. Oxford, UK: Oxford University Press. 2005b.

_____. **Roots and Categories**, handout of handout of keynote address to the Circle of Generative Grammar, University of the Basque Country, April 2009. Available: <http://webspace.qmul.ac.uk/hborer/downloads/roots_and_categories-revised.pdf>. Acess: 23 apr. 2014.

CHOMSKY, Noam. **Remarks on nominalizations.** In: Roderick A. Jacobs & Peter S. Rosenbaum (eds.). Readings in English Transformational Grammar. Waltham MA: ginn: 184-221. 1970.

MARANTZ, Alec. **No escape from syntax:** Don't try morphological analysis in the privacy of your own Lexicon. Proceedings of the 21st Annual Penn Linguistics Colloquium: Penn Working Papers in Linguistics. 1997.

PEDERNEIRA, Isabella L. **Etimologia e Reanálise de Palavras.** 2010. 123p. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2010.

LEMLE, Miriam e PEDERNEIRA, Isabella L. I**nserção lexical ou envoltório lexical?**. Alfa: Revista de Linguística *(UNESP)*, 2012. v. 56, n. 2, 469-490.