A Case Study of Copredication over a Deverbal that Reconciles Empirical Data with Computational Semantics

Livy Real, Christian Retoré

A Case Study of Copredication over a Deverbal that Reconciles Empirical Data with Computational Semantics

Livy Real[1] and Christian Retoré[2]

[1] IBM Research (São Paulo, Brazil)
[2] Université de Montpellier & LIRMM

Abstract. This paper is about the automated analysis of the lexical and compositional semantics of nominalisation, in particular of felicitous co-predications — infelicitous ones are rejected. We focus on Brazilian Portuguese nominalisation introduced by the suffix *-ura* but our discussion applies to other nominalisations and languages as well. Much of the theoretical work on deverbals, including ours, concluded that deverbal senses (process, result, location and so on) are rather unforeseeable from the verb and the suffix. The (in)felicity of co-predication, that is the possible conjunction of two predicates which applies to different senses, is also difficult to predict, and in the case of deverbals it is assumed to be (almost) impossible. We present here a study of the CHAVE corpus and show that CHAVE does actually contain some of the supposedly non-existent copredications. We explain our formalisation of sense variation and copredications in our logico-computational framework, Montagovean Generative Lexicon. We analyse the felicitous copredications and explain how the infelicitous ones are rejected. We give possible reasons for the felicity of the observed copredications, and finally we sketch out some guidelines for handling two notions that are necessary to a proper treatment of deverbals, namely local context and degree of semantic felicity.

1 Nominalisation and copredication

To account for the semantics of a compound expression, one needs to know which sense of its parts contributes, and how it contributes, to the meaning of the whole expression. For this aim, the predicates need to specify the nature of their arguments. For instance *to bark* applies to animals especially to dogs, while the following sentence is clearly infelicitous apart possibly in fictional contexts, as children stories.

(1) This chair barked.

Once a restricted and precise compositional framework is defined, meaning transfers and coercions require some flexibility in order to account of derived meaning, like:

(2) The head of department barked out orders to his secretary.

When a word has different meanings, an intriguing question is how do we deal with copredication, especially since the pioneering formal work of Pustejovsky [12].

The *copredications* on diverse senses of a given word is when we conjoin a predicate applied to one sense with another predicate applied to another sense, see next sentence:

(3) Your translation of Priscian's text, which has been revised many times, was placed on the desk.

Here one single token of the word *translation* refers both to the result of the act of translate and to its process.

Copredications are a challenge to the formal linguistic communities: their proper treatment, in a system that implements restrictions of selection, is difficult. Furthermore, as the senses themselves, they obey to language specific constraints that are possibly totally idiosyncratic and hard to learn from corpora — it is difficult to tell whether a copredication is rare or simply absent because it cannot be said. The computational semantics of co-predication, that is an automated semantic analysis which computes the semantics of meaningful sentences and rejects semantically ill-formed sentences, is quite difficult to achieve — let alone the difficulty to acquire the sophisticated lexicons that such an analyzer requires. Here are some examples of felicitous and infelicitous copredications:[3]

(4) Liverpool is a poor town.

(5) Liverpool is an important harbor.

(6) Liverpool is an important club.

(7) Liverpool is a poor town and an important harbor.

(8) * Liverpool is an important harbor and an important club as well.

Another issue that has been intensively discussed, both in purely linguistic and formal studies [6, 12], since at least [3], is the derivation of the sense of a *nominalisation*. Nominalisations are nouns derived from other part-of-speech words, as verbs (such nominalisations are called deverbals), adjectives or even nouns. How does one derive the possible senses of a nominalisation from the sense of the initial word? It seems there is not final answer to it, specially because many nominalisations have idiosyncratic meanings, as *estacionamento* ('parking') formed by *estacionar* ('to park') + *-mento*. The suffix *-mento* forms eventive nouns, but *estacionamento* has only the locative reading and not the eventive one and this cannot be compositionally predicted.

Nominalisations have been intensively studied in linguistics, but also in NLP and knowledge representation. Some rules managing their behavior have been proposed, but nothing fully satisfying appeared. Indeed, nominalisations cause a problem to every lexical theory that understands words (or parts of words, or lexicalised expressions) as static parts of discourse. Sometimes nominalisations behave as common nouns (without any special 'deep' structure), sometimes they seem to require specific arguments (and contexts) to achieve a given sense. According to some studies, as [6], those arguments correspond to the argument structure of the associated verb, but many lexical theories cannot account for this morpho-semantic relation (see [13] for a deeper discussion).

[3] A "*" indicates a bad sentence, either semantically of syntactically.

Moreover, many examples indicate that the relationship between a verb and its nominalisation is not necessarily based on the argument structure of the verb, it may involve circumstantial complements as well, as noted by [15]. This error comes from studies that only consider prototypical nominals (as *construction*, *destruction* and *translation*) with a set of possible senses restricted to process and result, which does not account for the many difference senses a nominalisation can have. Of course one may infer from corpora or lexical networks *some* information about the deverbal meaning but computing the lexical sense of the deverbal needed for compositional semantics from the one of the associated verb and the suffix is presently out of reach. Derivational relations, that are morphological relations that link a noun-verb pair can be extracted but extracting semantic information is much harder. Most of lexical resources do not have this kind of semantic information encoded as it has to be done manually and it is a very time consuming task that also needs to be done by experts. However, biggest resources have been encoding the semantic type of a nominalisation manually. In [4], Princeton Wordnet team explains how they encoded semantic information about deverbals: starting from a previous list of morphological related items, they manually checked each pair to introduce in Princeton's Wordnet morpho-semantic links, that relate verb-noun pairs through typed relations, as agent, event and instrument.

To close this section, we present below two examples of copredications involving three senses of the Portuguese deverbal *assinatura* ('signature, signing') — i. discussion leading to an agreement, ii. writing act, iii. result all of them as a grapheme:

(9) * A assinatura atrasou três dias e estava ilegível.
The signature took three days and was illegible.

(10) A assinatura furou a folha e estava ilegível.
The signature pierced the sheet of paper and was illegible.

2 Our computational-logic framework

We use a logical framework known as Montagovian Generative Lexicon (MGL), deeply presented in [16] and firstly introduced in [8]. Inspired by Pustejovsky's Generative Lexicon [12], MGL extends Montague semantics and can also be viewed as an extension of Muskens' Ty_n, so it is also called as MGL ΛTy_n. As a formal definition of MGL would be too lengthy, we present here informally its principles, stressing the similarities and differences with related systems.

As Montague's initial work [9], this system is closely related to **syntax**, in particular to categorial grammar. If the semantic lexicon is known, i.e. if the association of typed terms to words is known, as in [9], then the semantic representation can be automatically computed. First one substitutes in the parse tree the lexical lambda terms for the words, then one obtains a properly typed term, which possibly makes use of the optional terms. After reduction, this yields a logical formula of higher order predicate calculus (or nothing: if the type mismatch is not resolved, it is an infelicitous sentence). The reduction, which is a strongly normalizing process, guarantees that the semantic interpretation of a sentence always is a logical formula, i.e. something that makes sense.

There are no specific rules for combining those terms. Only applications and abstractions rules are needed — there can be some abstraction steps, because Lambek syntactic parse trees may contain some abstractions.

The semantic part of the lexicon associates each word with a main typed lambda term (the usual Montague lambda term) and several optional typed lambda terms (including at least the identity of the type of the main term). These typed terms are in system F: basically F is an easy generalization of simply typed lambda calculus which allows to factorise terms that act uniformly.[4] Here are some more details on types and terms:

- Base types includes t (type of propositions, that we would better call *prop*) but in addition to Montague there are many base types for entities e_i that can be considered as sorts of a many sorted logic.
- In addition to base types, we have type variables, and in addition to functional types, we have quantified types obtained by quantification over a type variable, i.e. we are using second order typed lambda calculus (Girard's system F, introduced in [5]).
- Quantification over type variables is used to factorize operations like conjunction, or for determiners and quantifiers that acts uniformly over types — for instance the conjunction "and" used for copredication over a sense of a word w of a type B and another sense of the word w of another type A with two predicates of respective types $a \to t$ and $b \to t$ will apply to all types a and b, hence will contain a quantification over all types. But having many simply typed terms, e.g. for the conjunction used for copredication, would work just the same. [5]
- Reduction is defined as expected:

  $$(\lambda x^U t^Y)(u^U) \longrightarrow t[x^U := u^U] \quad u \text{ is a term of type } U \text{ and } x \text{ a variable of type } U$$

  $$(\Lambda X t)\{T\} \longrightarrow t[X := T] \quad X \text{ is a type variable and } U \text{ is a type}$$

  An important property is that reduction is a strongly normalising process, which preserves the types. Thus, the semantic lambda term associated with a sentence of type t reduces to a normal term of type t: it is easily seen (see e.g. [10, chap 3]) that this normal term, which only involves logical constants and predicates, corresponds to a logical formula.

Intuitively the optional lambda terms, which have a functional type, allow to view a word w of type A as a word of type B hence they give access to a sense of w different from the initial one that is described by the type of the main lambda term of w.

Semantic infelicity is described by type mismatch, i.e. when a functional term of type $A \to X$ is applied to a term t of type B with $B \neq A$.

The optional terms that the lexicon associate with the words of the analysed phrase are used to fix such type mismatches. To fix a type mismatch as above, the lexical entry of one of the words should have an optional term of type $B \to A$.

[4] Let us take a programming example: a sorting function, like QuickSort, applies to a list of A object, and an order on A object and yields a sorted list of A objects: you can factor such a program/term over all types A, because it works just the same on every type.

[5] We also have subtyping, in particular on the sorts e_i, but we do not focus on it here.

Incompatibilities are accounted for as folllows. When one sense associated with a word (that corresponds either to the main lambda term or to an optional lambda term) is incompatible with any other, the optional lambda term corresponding to this sense (or the identity type in the case of the main lambda term) is declared in the lexicon to be *rigid* — otherwise the term is said to be *flexible*.

Quantification over types allows a factored treatment of conjunction which is extremely useful for modeling copredications, as discussed in [15, Section 4.1] or [16]: *whenever* an object x of type ξ can be viewed both as an object of type α to which a property $P^{\alpha\to\mathrm{t}}$ applies and as an object of type β to which a property $Q^{\beta\to\mathrm{t}}$ applies (via two terms $f_0 : \xi \to \alpha$ and $g_0 : \xi \to \beta$), the fact that x enjoys $P \wedge Q$ can be expressed by the unique polymorphic term:

$$\&^{\pi} = \Lambda\alpha\Lambda\beta\lambda P^{\alpha\to\mathrm{t}}\lambda Q^{\beta\to\mathrm{t}}\Lambda\xi\lambda x^{\xi}\lambda f^{\xi\to\alpha}\lambda g^{\xi\to\beta}. \\ (\&^{\mathrm{t}\to\mathrm{t}\to\mathrm{t}}\ (P\ (f\ x))(Q\ (g\ x)))$$

Using F or any other Type Theory, as Martin Löf's theory, does not make a big difference: the impredicative quantification instead of predicative quantification does not make a big difference. The difference rather lies in the organization of the lexicon with optional lambda terms. The semantic representations are formulae of higher order many sorted logic and not type theoretical formulae — the later ones cannot be interpreted in usual set theoretic models.

3 A typical case: *-ura* in Brazilian Portuguese

In some previous papers [15], we studied the nominalizations in Brazilian Portuguese (BP from now on) both formally and empirically and especially the ones that are constructed with the *-ura* suffix as *abertura* 'opening', *assinatura* 'signing/signature' and *brancura* 'whiteness'. The choice of the suffix *-ura* was made for many reasons: this work can certainly be used to understand other similar nominalisation suffixes at least in other neo-Latin languages as *-ure* in French (*coupure* 'cut'), *-ura* in Catalan (*obertura* 'opening') and *-ura* Spanish (*blancura*, 'whiteness') and possibly be extended to not so close languages, as German and its suffix *-ung* and *-ní* in Czech.

Our descriptive and theoretical previous work on these deverbals led us to the following conclusions:

- From a corpus based analysis [14], we provided 13 types that *-ura* can mean: event, result, physical result, locative, collective, means, property, instrument, a given portion, rest, function, duration of a function, science/art. From theoretical words, it was known to have at least eight senses ([19, 18, 11]), including eventive, resultative, locative and collective readings, just as other well-known morphemes like *-ung* in German and *-age* in French.
- It seems hopeless to predict the senses of *-ura* nominalisations, as well as the compatibility of these sense, at least for the time being no one knows how they could be acquired from corpora. Observe that senses are listed in a dictionary, while their pairwise compatibility is not.

- When the senses and their compatibility w.r.t. copredication are known, i.e. when someone can write the lexical entries, MGL faithfully models the deverbal, i.e. computes exactly the correct semantic representation(s), in particular it accepts exactly the possible copredications – and rejects the ones that are semantically ill-formed.
- Before the present corpus study, we concluded from examples in the literature that copredication was almost always impossible on the different senses of a deverbal, a possible exception being example (10) above.

Although the literature and our own work agreed that copredication on the different senses of a deverbal are almost impossible, we started to have doubts about this claim and wanted to confront it to corpus study. Indeed, while corpora cannot prove that a construction does not exist, they can prove that a construction exists, by providing meaningful examples of the construction.

As far as we know, [14] was the first work that considers nominalisations in corpora, at least for Portuguese. It offers an enriched description of a relevant fragment of deverbal nominals in BP, considering all possible senses of each nominalisation, all possible co-predications over those senses and a large corpora analysis based on Corpus Brasileiro[6] that did not find any co-predication with words formed by *-ura*. Therefore we generated a large list of possible co-predications and tested their acceptability with three native speakers of (Brazilian) Portuguese and most of the sentences were rejected by the three of them. The acceptable sentences were often using very frequent words of the common Brazilian lexicon. Thus [14] claimed that the acceptability of co-predication structures is correlated to the frequency of the considered *-ura* word. Although this work provided some results, we decided to search for co-predication with words formed by *-ura* in another corpus keeping in mind the influence of the frequency on the felicity of the copredication. Indeed, the corpus Brasileiro is not a very reliable corpus[7], it has much noise and does not seem to be correctly tagged.

4 A closer look at data

Thus, we searched for copredications involving *-ura* words in another corpus, we chose the CHAVE corpus[8], which contains news from two relevant newspapers of Portugal (*Publico*) and of Brazil (*Folha de São Paulo*) from 1994 and 1995, with almost 1 million words.

Our search looked for the syntactic structures that may produce acceptable copredications, namely the ones of following the criterion given in [7] [9]:

[6] Corpus developed by Tony Berber Sardinha & alii, trough AC/DC tool [20]. Freely available on `http://www.linguateca.pt/ACDC`.

[7] We would like to thank Claudia Freitas for the discussions on corpora choice and the help with the search engines of Linguateca corpora.

[8] Freely available on `http://www.linguateca.pt/acesso/corpus.php?corpus=CHAVE`.

[9] As shown in [15], this criterion is neither sufficient nor necessary for the felicity of copredications, but we thought that copredications following these patterns are more likely to occur.

Syntactic constraints on co-predication on different meanings of a nominalisation:

i. *Split co-predication between main clause and subordinate clause;*
   - *The construction took ages and is of Victorian style. (non split)*
   - *The construction which took ages is of Victorian style. (split)*
   - *The construction which is of Victorian style took ages. (split)*

ii. *Temporal disjunction between the two predications;*
   - *The construction is taking ages and is of Victorian style. (no temporal disjunction)*
   - *The construction is taking ages and will be of Victorian style. (with temporal disjunction)*

iii. *Omission of the internal argument.*
   - *The construction took ages and is of Victorian style. (without internal argument)*
   - *The construction of the house took ages and is of Victorian style. (with internal argument)*

Here we bring an example from [7] of a sentence that respects those constrains:

(11) La costruzione, che si protrasse fino al XVII secolo, rimane un'importante testimonianza della geniale tematica del Palladio.

'The building, which continued till the XVII century, represents an important evidence of Palladio's ingenious artwork.'

Above we have two split clauses that point to different periods of time and a possible internal argument of 'building' is missing.

Since the main structure of subordinate clauses in BP is formed by ', que' (*, that*), we captured sentences with main clause and subordinate clause (unfortunately, there was no way to search in the corpora temporal disjunction) by the formal expression:

```
[lema="-ura word"][func="N<.*"]*\,[pos="N"]*"," "que"
```

From our previous work [14], we have found that the frequency of occurrence of a given word in the lexicon can have an important role on co-predication acceptability. So, from a list of all *-ura* nominalizations in Portuguese, obtained on OpenWordnet-PT [10], we separated all the nominalizations into three blocks: very frequent, frequent and rare words, considering the total number of occurrences of each word in CHAVE. Then we arbitrarily picked 2 words from each of these groups and searched on CHAVE corpus for copredications with them[11].

[10] Freely available on `http://wnpt.brlcloud.com/wn/`.

[11] *Abertura* 'opening', *assinatura* 'signing/signature', *arquitetura* 'architecture', *legislatura* 'legislature', *moedura* 'milling', 'textitmoedura 'rinse'.

W O R D	# occurrences in CHAVE	# occurrences in possible co-predication structures
abertura	12900	221
assinatura	4194	152
arquitetura	1105	28
legislatura	990	46
moedura	1	0
enxaguadura	0	0

From the previous expression, we found 437 sentences which we manually analysed searching for co-predications, and 4 of them can be considered co-predications:

(12) *Physical result (φ)+ event(v)*
Por outro lado e em um domínio totalmente diferente, a **assinatura**$^{\varphi \wedge v}$ que [ele traça a partir do próprio nome]$^{\varphi \to t}$ [coroa a chegada do patrimônio]$^{v \to t}$.

On the other hand and in a totally different field, the **signature**$^{\varphi \wedge v}$ that [he draws from his own name]$^{\varphi \to t}$ [crowns the arrival of heritage]$^{v \to t}$.

(13) *Physical Result + Informational Content*
Retomando o texto, mostra como [várias entidades se organizaram e recolheram]$^{\varphi \to t}$ em momentos diferentes milhares de **assinaturas**$^{\varphi \wedge info}$ que [reforçaram e respaldaram as conquistas da população contra a empresa poluidora]$^{info \to t}$.

Returning to the text, it shows how [various entities organized themselves and collected]$^{\varphi \to t}$ at different times thousands of **signatures**$^{\varphi \wedge info}$ which have [strengthened and endorsed the achievements of the population against the polluting company]$^{info \to t}$.

(14) *Result + Event*
As peças agora no Museu Pushkin, em Moscou, e antes no Museu de Pré-História e História Antiga de Berlim, foram achadas na altura de Tróia 2, uma cidade com **arquitetura**$^{v \wedge r}$ [imponente]$^{r \to e}$ que [floresceu de 2.500 a 2.200 a.C.]$^{v \to t}$, na Idade do Bronze.

The pieces now in the Pushkin Museum in Moscow, and before the Prehistory and Ancient History Museum of Berlin, were found at the time of Troy 2, a town with an [imposing]$^{r \to e}$ **architecture**$^{v \wedge r}$ that [flourished 2500-2200 BC]$^{v \to t}$ in the Bronze Age.

(15) *Function Duration + Function*
Uma **legislatura**$^{dur \wedge func}$ que [começa ao mesmo tempo em que começa o mandato de um presidente eleito provavelmente em dois turnos]$^{dur \to t}$ [deverá aguardar o envio das principais mensagens do presidente]$^{func \to t}$, que consubstanciem o seu programa de governo.

A **legislature**$^{dur \wedge func}$ that [starts at the same time that a president probably gets elected in two rounds]$^{dur \to t}$ [should wait for the key messages of the president to be sent]$^{func \to t}$, in order to substantiate the government program.

From 437 found sentences, 4 copredications represent 0,9% which is a high number if we consider that this kind of morpho-syntatic construction is expected to not normally happen.

Let us take 12 as an example. In 12, 'assinatura' ('signing/signature') in *a assinatura que ele traça a partir do próprio nome* 'the signature that he draws from his own name" has the physical result reading, as it means the mark on the paper that comes from the signing event. Therefore 'assinatura' in *a assinatura coroa a chegada do patrimônio* '[the signing] crowns the arrival of heritage' is related to the event of signing: it is not the mark on the paper that 'crowns the arrival of heritage', but the fact that the document was signed. It is interesting to note that the same lexical item in Portuguese that concomitantly belongs to the type event and to the type physical result is expressed in English by two different words: 'signature' (the physical result) and 'signing' (the event itself).

Here we consider the follow structure as the lexical entry for *assinatura* (that actually can hold much more types than these two types):

word	main λ-term	optional λ-terms	
assinatura	$\lambda x^{v}.sig(x) : v \to t$	$Id_{v} : v \to v$	$(flexible)$
		$f_{\varphi} : v \to \varphi$	$(flexible)$

Fig. 1. The lexical entry for *assinatura*

In Figure 1 the base types are defined as follows: v (events) and φ (physical objects).

(16) the main term is $\lambda x^{v}.(sig^{v\to t} x)$

(17) and the optional morphisms are

a. $Id_{v} = \lambda x^{v}.x$, the (always present) identity (referring to the agreement process) which is declared as *rigid*,

b. $f_{\varphi}^{v\to\varphi}$ turning the event into a physical object and declared to be flexible.

For a semantic analysis of *assinatura* in those two clauses, the first step is the composition of *assinatura* with the definite article *a* ('the'), which we deeply explain in [15, 17]. Here we assume $(sig)^{v}$ as the formula of *a assinatura*, considering *assinatura* applied to a/the = $\iota \lambda \alpha.(\alpha \to t) \to \alpha$.

Let us recall the polymorphic conjunction from section 2:

$$\&^{\Pi} = \Lambda\alpha\Lambda\beta\lambda P^{\alpha\to t}\lambda Q^{\beta\to t}\Lambda\xi\lambda x^{\xi}\lambda f^{\xi\to\alpha}\lambda g^{\xi\to\beta}.(t^{t\to t\to t}\ (P\ (f\ x))(Q\ (g\ x)))$$

Briefly, variable Q and P are the predicates to be copredicated, which apply to objects of types α and β respectively. Objects of type ξ can be mapped via f and g into α and β objects respectively ("viewed as"). The following term is able in any such situation to conjoin two different predicates that respectively apply to two different senses of the same lexical entry, if this lexical entry has a function (f and g) that changes its previous type (α and β) to the required type (ξ). An easy example is *a book* which can be viewed as a physical object and as an informational content. Therefore *a book* can be both *heavy* and *interesting*, as the predicate *heavy* applies to the *physical* sense of book and the predicate *interesting* applies to the *informational* sense.

The instantiations for our example should be as follows:

- $\alpha = v$, $P = crown^{v \to t}$, $f = f_v$,
- $\beta = \varphi$, $Q = draw^{\varphi \to t}$, $g = f_\varphi$,
- $\xi = v$, $x = sig^v$.

where:
sig^v = a assinatura
$crown^{v \to t}$ = *[a assinatura] coroa a chegada do patrimônio* '[the signing] crowns the arrival of heritage'
$draw^{\varphi \to t}$ = *traça [a assinatura] a partir do próprio nome* 'he draws [the signature] from his own name'

The polymorphic "and" $\&^{\Pi}$ takes as arguments two properties P (here: $crown^{v \to t}$) and Q (here: $draw^{\varphi \to t}$) of entities of respective type α (here: v) and β (here: ϕ), and returns a predicate that applies to a term x of type ξ. This predicate says that:

- if x of type ξ (here sig^v of type v):
  - enjoys P (here $crown(f_v(x))$) when viewed as an object of type α (here v) via f_v
  - enjoys Q (here $draw(f_\varphi(x))$) when viewed as an object of type β (here φ) via g (here f_φ)
- then x endowed with the proper meanings has both properties that is

$$crown(f_v(x)) \;\&^{\Pi}\; draw(f_\varphi(x))$$

Hence the copredication in example (12) can be derived.

Observe that if (as wrongly asserted in some papers) the senses of *assinatura* were incompatible, we could not derived this copredication. If they were incompatible then, in our semantic lexicon one of the two morphisms (identity and the function that maps the main sense to physical object sense) would be declared to be rigid. Here they are both flexible hence can be used both, as we did to derive the meaning of the copredication.

5 Consequences and ongoing work

These examples show that the copredications between different senses of a deverbal are not impossible as they are said to be in theoretical studies and in our previous corpus based analysis. Actually, our previous work considers only Brazilian Portuguese and CHAVE corpus is composed by European and Brazilian Portuguese, which leaves us the task of reproducing this experiment with two very reliable corpora of different kinds of Portuguese to check if there is a language specific constrain on copredications.

Our previous suggestion that the role of word frequency in the lexicon is also relevant to the felicity of copredications was not entirely proved, as from the six chosen words we work with, the most frequent word, *abertura*, does not appear in any copredication structure. However we do not consider this present work evidence enough to discard this suggestion and we still believe that the frequency of occurrence of a given word may play a role in copredication constructions.

Our MGL formalisation and computational analysis reject the infelicitous copredications and compute the semantics of felicitous copredications (just as we analysed the homemade and slightly far-fetched example (10) or rejected (9) in [15]).

The examples above of correct copredications suggest some hints on the semantic conditions that make copredications possible (in addition to the syntactic constraints reminded above). A likely semantic explanation, which is in the air at least for ordinary common nouns like institutions and the which we would like to confirm, is that the copredications are more felicitous when either:

1. The two predications belong to the same domain which gathers the two senses.

   (18) Barcelona organised the Olympiads and won four Champions Leagues. (institution/ football club)

2. There is an opposition between two predications.

   (19) Huelva, a small southern town, defeated Barcelona. (institution/ football club)

Hence this possibility depends on the local context and this probably can be related to the local coercions discussed in [1]. Here we have local compatibilities, and once more for semantics the (local) context plays a central role. However since the relation between this model is up to now unrelated to syntax or discourse structure, it is presently hard to tell what is exactly the domain of locality of local coercions.

We can try to explain the felicity of the copredications over deverbals that we found in the CHAVE corpus by means of the two principles suggested above. This is just a try, which deserves to be discussed from data and psycholinguistic experiments. In the first CHAVE example, (12), the notion of achievement, supported by the word *coroa* ('crowns'), gathers in the same context both senses, i. resulting grapheme and ii. agreement process, that are used by the two predicates. In the second CHAVE example (13), both senses of *assinatura*, as a grapheme and as the informational content expressed by the grapheme, share the notion of engagement (to defend the text). In the third CHAVE example (14), the two aspects, result and process (brief, when compared to the time that elapsed since then) can be somehow identified when viewed from several thousand years later, the common feature being the notion of an instantaneous event, sense which is supported by the verb *floresceu*. In the fourth and last CHAVE example (15), both predicates speak about a duration, since the people who should be waiting are precisely the ones elected for this duration. In the example (10), the opposition between the time to write the signature and its illegibility makes the copredication works, since world knowledge makes us to expect that things which take a long time to be made are good/useful.

As we may observe from these last examples or from example (10), the semantic felicity is not something that either holds or not. Hence a scalable notion of acceptance/felicity would be much welcome. For type theoretical framework like MGL and others, scalable type judgment (semantic representation of type t) as initiated by [2] should be relevant. This scalability question is somehow orthogonal to the precise framework and therefore should be easy to implement for MGL once scales are integrated to type judgment.

We are presently exploring these possibilities.

References

1. Nicholas Asher and Zhaohui Luo. Formalization of coercions in lexical semantics. In Emmanuel Chemla, Vincent Homer, and Grégoire Winterstein, editors, *Sinn und Bedeutung 17*, pages 63–80, 2012.
2. Robin Cooper, Simon Dobnik, Shalom Lappin, and Staffan Larsson. A probabilistic rich type theory for semantic interpretation. In *Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics (TTNLS)*, pages 72–79. Association for Computational Linguistics, 2014.
3. D.A. Cruse. *Lexical semantics*. Cambridge textbooks in linguistics. Cambridge University Press, 1986.
4. Osherson A. Fellbaum, C. and P.E. Clark. Putting semantics into wordnet's m̈orphosemanticl̈inks. *Proceedings of the Third Language and Technology Conference, Poznan, Poland. Reprinted in: Responding to Information Society Challenges: New Advances in Human Language Technologies, eds. Z. Vetulani and H. Uszkoreit. Springer Lecture Notes in Informatics vol. 5603:350-358 (2009)*, 2007.
5. Jean-Yves Girard. Une extension de l'interprétation de Gödel à l'analyse et son application: l'élimination des coupures dans l'analyse et la théorie des types. In Jens Erik Fenstad, editor, *Proceedings of the Second Scandinavian Logic Symposium*, volume 63 of *Studies in Logic and the Foundations of Mathematics*, pages 63–92, Amsterdam, 1971. North Holland.
6. J. Grimshaw. *Argument Structure*. MIT Press, 1990.
7. Elisabetta Ježek and Chiara Melloni. Nominals, Polysemy, and Co-predication. *Journal of Cognitive Science*, 22:1–31, 2011.
8. Bruno Mery, Christian Bassac, and Christian Retoré. A Montagovian generative lexicon. In *Formal Grammar*. CSLI, 2007.
9. Richard Montague. English as a formal language. In Bruno Visentini, editor, *Linguaggi nella Societa e nella Tecnica*, pages 189–224. Edizioni di Communità, Milan, Italy, 1970. (Reprinted in R. Thomason (ed) *The collected papers of Richard Montague* Yale University Press, 1974.).
10. Richard Moot and Christian Retoré. *The logic of categorial grammars: a deductive account of natural language syntax and semantics*, volume 6850 of *LNCS*. Springer, 2012.
11. A. Nonymous. This is one the papers involving at least one of the authors. It is skipped to respect the anonimity of the submission. If the paper is accepted the reference will be given in the final version.
12. James Pustejovsky. *The generative lexicon*. M.I.T. Press, 1995.
13. Livy Real. *Nominalizações*. PhD thesis, Universidade Federal do Paraná, Curitiba, April 2014.
14. Livy Real and Alexandre Rademaker. An overview on portuguese nominalisation. In *Workshop on TYpe Theory and LExical Semantics*, Barcelona, aug 2015.
15. Livy Real and Christian Retoré. Deverbal semantics and the Montagovian generative lexicon $\Lambda\mathsf{Ty}_n$. *Journal of Logic, Language and Information*, pages 1–20, 2014.

16. Christian Retoré. The Montagovian Generative Lexicon ΛTy_n: a Type Theoretical Framework for Natural Language Semantics. In Ralph Matthes and Aleksy Schubert, editors, *19th International Conference on Types for Proofs and Programs (TYPES 2013)*, volume 26 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 202–229, Dagstuhl, Germany, 2014. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
17. Christian Retoré. Typed Hilbert epsilon operators and the semantics of determiner phrases (invited lecture). In Glyn Morrill, Reinhard Muskens, Rainer Osswald, and Frank Richter, editors, *Proceedings of Formal Grammar 2014*, number 8612 in LNCS/FoLLI, pages 15–33. Springer, 2014.
18. Luiz Carlos de Assis Rocha. A nominalização no português do brasil. *Revista de Estudos da Linguagem*, 8(17), 1999.
19. Antônio José Sandmann. *Formação de palavras no português brasileiro contemporâneo*. Scientia et Labor, 1989.
20. Diana Santos and Luís Sarmento. O projecto ac/dc: acesso a corpora/disponibilização de corpora. In Amália Mendes and Tiago Freitas, editors, *Actas do XVIII Encontro Nacional da Associação Portuguesa de Linguística (APL 2002)*, pages 705–717, Lisboa, 2003. Associação Portuguesa de Linguística APL.