

PrepNet.Br: a Semantic Network for Prepositions

Déborá D. Garcia¹, Bento Carlos Dias da Silva²

¹Programa de Pós-Graduação em Linguística (PPGL)
Universidade Federal de São Carlos (UFSCar) - São Carlos, SP – Brazil

²Departamento de Letras Modernas (DLM)
Universidade Estadual Paulista (UNESP) - Araraquara, SP – Brazil

{deboradom@gmail.com, bento.silva@gmail.com}

***Abstract.** This article presents the idea of the PrepNet.Br, a semantic network consisting of synsets of prepositions to be built for Brazilian Portuguese along the lines of the French PrepNet [Saint-Dizier 2005]. This enterprise is of relevance to the linguistic description of the class of prepositions as well as to Natural Language Processing resource-building. The construction of this network also brings a linguistic analysis alternative to the traditional one, investigating the polysemic nature of prepositions within a cognitive view of language.*

1. Introduction

In the last decades, many studies have been made about nouns, verbs and adjectives, both in Linguistics and Computational Linguistics. The study of prepositions, however, has been more modest due to its high degree of polysemy, which makes it difficult to predict their semantics and their realizations across different languages [Saint-Dizier 2006]. For example, while in Portuguese the same preposition can occur in three distinct contexts - Você mora *no* (em+o) campo; Você conhece pessoas *na* (em+a) festa; and Você entra *em* férias -, in English, we have three different prepositions (*in*, *at* and *on*) in the very same contexts - You live *in* the country; You meet people *at* the party; and You go *on* holiday” [Taylor 1995], respectively.

An accurate model of preposition usage is crucial to avoid repeatedly making errors in terms of parsing and generation. This paper outlines the idea of a semantic network of prepositions to be built for Brazilian Portuguese – the PrepNet.Br – that aims to do a better characterization of this part of speech.

2. Prepositions and different approaches

In previous studies about prepositions, the biggest challenge has always been to define its semantic value, since there is a tradition that excludes or limits the inherent significance of these items to the syntax.

For example, traditional grammars usually describe prepositions as grammatical items that receive some kind of meaning only in context. This perspective assumes that the native speaker must master the prepositions one by one, because their use is idiomatic and therefore "must be memorized". This results in an unsystematic characterization of prepositions, in which there is no agreement regarding its semantics.

However, the facility that the native speakers learn how to use prepositions in every new context indicates that its usage is highly structured by human mind and the alleged flexibility in the use of prepositions would be guided by some (mental) logic, not just memorization.

As a consequence of this perception, there are new studies that look to the plurality of meanings that each preposition takes in different contexts from a semantic point of view. To the cognitive framework, for example, prepositions do have a certain kind of meaning – their meaning is probably more complex than the meanings of other lexical categories – and the many uses of a preposition are considered “extensions of its core meaning”. Thus, prepositions have been discussed as polysemous items. [Castilho 2010].

The cognitive framework supports the idea of a lexicon-grammar *continuum*, where “all the prepositions, regardless of their degree of grammaticalization, can introduce adjuncts, expressing various relations and senses”¹ [Ilari et.al. 2008]. For example, the prepositions “de” (Portuguese) and “on” (English), instantiate either grammatical items (*Eu dependo de você / I depend on you*) or lexical items, which bear semantics (*Maria é de São Paulo / Maria is from São Paulo* (Source), and *O jornal está sobre o tapete / The newspaper is on the mat* (Spatial Location)).

Once presented the view that prepositions are not only grammatical items, we introduce the idea of the PrepNet.Br: a semantic network of prepositions whose structure aims to represent their meanings and usage in terms of a particular computational linguistic representation.

3. A network of prepositions

From a technological point of view, prepositions have great importance to enrich and assist Natural Language Processing (NLP) applications, since they encode essential meanings for understanding the propositions (the logical and conceptual meanings of sentences). For example, prepositions conceptualize **location** (*put the book on the shelf*), **instrumentality** (*cut the meat with a knife*), **origin-goal** (*traveled from São Paulo to Rio de Janeiro*), **recipient** (*the wine given to his friend*), **time interval** (*arrive between noon and 1 p.m.*) and **space location** (*it was between the table and the wall*). These concepts are vital to precise language understanding.

A PrepNet network is a computational linguistic resource for NLP stemming from Saint-Dizier (2005, 2006): a repository of prepositions from different languages with a formal specification of their syntactic and semantic behaviors. This idea has three motivations: (i) to construct a system similar to *wordnets* [Miller; Fellbaum 1991] and with the possibility to complement them; (ii) to assign thematic roles [Bakker; Siewierska 2002; Jackendoff 1991], and, above all, (iii) to reach a more complete and robust conceptual description of prepositions. Saint-Dizier (2005) believes that a PrepNet should be the starting point for a better characterization of prepositions, necessary before analyzing their interaction with verbs, for example.

¹ Original version (portuguese): “todas as preposições, independentemente de seu grau de gramaticalização, podem funcionar como introdutoras de adjuntos, expressando as relações e os sentidos mais variados”.

In a PrepNet, the different uses of a preposition should be organized around a small number of general senses. That is, the meaning of each preposition is analyzed in terms of its primary (core) sense, which may be the source for further analysis in terms of metaphorical senses. It is known that there is no full synonymy between two or more words, and therefore there is no complete overlap between two or more preposition meanings. However, when we consider that each preposition presents a set of semantic possibilities, partial coincidences between them are possible [Borba 1971]. Therefore, it is quite feasible to organize prepositions in terms of synsets (sets of cognitive synonyms), *à la* wordnets.

In his papers, Saint-Dizier (2005, 2006) presents an initial classification for French prepositions, however it seems not to have progressed beyond the preliminary stages. Inspired by his idea and taking the liberty to carry out formal and methodological changes, we propose the PrepNet.Br. Figure 1 illustrates the analysis we are working on: the description of the contents and layout of Portuguese (Brazil) and English preposition synsets, which provides their grammar and semantics, and the alignment of the “equivalent synsets” of both languages.

	Language: Portuguese (Brazil)	Language: English	
Linguistic Level	Synset: {a1, para, em1}	↔	Synset: {to1}
	Sample sentences Ele foi levado para a cela individual. Só podia sair para ir ao banheiro. (...) então a gente vai no chá, neh?		Sample sentences He went to the shop. He walked to the house. He goes to school at eight o'clock.
Conceptual-Semantic Level	Gloss: “spatial direction” Image Schema: DYNAMIC PATH Schema (“goal”): The FIGURE is at the final location of a path. Spatial Axis: Horizontal Semantic Feature: /GOAL/ Family: Localization Facet: Goal Modality: ? Evoked Frame: <u>Goal</u> : “A Landmark (in combination with the image schema evoked by particular targets) serves to pick out the final location of a Trajector in a construed or actual motion event.” Inherits from: Locative_relation, Trajector-Landmark Used by: Source_Path_Goal		

Figure 1 – A preliminary version of a PrepNet.Br synset and its English equivalent.

From Saint-Dizier’s proposal, we kept (I) the essential notion of synset (enriched by short definitions and sample sentences illustrating the use of the synset members); (II) the description of prepositional senses in two levels (linguistic and conceptual); and (III) the characterization of semantic features in terms of “semantic family”, “facets” and “modalities of a facet” (three levels proposed by the author). The changes proposed for the PrepNet.Br were the inclusion of (i) the semantic-conceptual classification of

prepositions presented in Ilari et al (2008), in terms of Image Schema and Spatial Axes; and (ii) *frames* from FrameNet [Fillmore 1982; Ruppenhofer 2010].²

The organization of prepositions in terms of synsets is backed up by the hypothesis that prepositions that share semantic features, i.e. share the same meaning in certain contexts, are synonymous. The semantic consistency of a synset is guaranteed by checking whether their constituent prepositions are interchangeable in sample sentences, assuming that the user's intuition guides the validation.

Finally, we outline the strategy to discover and align Brazilian Portuguese and English synsets. When English prepositions express a meaning specified by a synset of Brazilian Portuguese prepositions, we verify if those prepositions are interchangeable in different sentences taken from corpora, and then a synset with these English prepositions is proposed. In Figure 1 it was possible to construct the English synset {*to1*} and align it to the Portuguese synset {*a1, para, em*}, since the prepositions *to1* can be replaced in the English sentence “I take them **to** school and go to work”, which is equivalent to the Brazilian Portuguese sentence “Eu os levo **para** a escola e vou trabalhar”. However, care must be taken with cases such as this: the pair of sentences “Temos que escolher **entre duas** alternativas”/“We must choose **between two** alternatives” and “Temos que escolher **entre três** alternativas”/“We must choose **among three** alternatives” reveals two possibilities worth exploring, either one synset for Brazilian Portuguese and two synsets for English ({*entre1*} = {*among*}; {*between*}), where {*entre1*} is considered a generalization of both English synsets or two distinct synsets for both languages ({*entre1.1*}={*between*}, which is used in speaking of two discrete entities, and {*entre1.2*}={*among*}), which is used in speaking of three or more entities collectively. A choice that calls for deeper analysis.

4. Final words

With those descriptive and analytical elements, Garcia (2013) carried out the analysis of a set of Brazilian Portuguese spatial prepositions and the systematization of their characteristics in terms of synsets, resulted in the construction of 13 synsets aligned to 14 English synsets. At this first stage, prepositional phrases and metaphorical senses were set aside.

By presenting the conception of a PrepNet.Br, this short paper aims to contribute to the first steps towards a more accurate computational-linguistic description of prepositions and to launch out the idea of a prepositional network to add to Brazilian Portuguese NLP resources.

²It is worth mentioning that including the description of preposition semantics in terms of frames allows further structural and conceptual connections between PrepNet and FrameNets.

References

- Bakker, D.; Siewierska, A. (2002) Adpositions, the lexicon and expression rules. In Mairal Usón, R.; Perez Quintero, M. J. (editors), *New perspectives on argument structure in functional grammar*. Berlin, Mouton de Gruyter (pp. 125-77).
- Borba, F. S. (1971) *Sistemas de preposições em português*. PhD Thesis. Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 1971.
- Castilho, A. T. (2010) *Nova gramática do português brasileiro*. São Paulo, Contexto.
- Fillmore, C.J. (1982) Frame semantics. In *Linguistics in the Morning Calm*. Seoul: South Korea: Hanshin Publishing Co., (pp. 111 -137).
- Garcia, D. D. (2013). *Construção exploratória de uma PrepNet para o português do Brasil: uma incursão linguístico-computacional no universo das preposições indicativas de espaço*. Master Thesis. Faculdade de Ciência e Letras, Universidade Estadual Paulista.
- Ilari, R.; et al. (2008) As preposições. In Ilari, R.; Neves, M. H. M. (Orgs.) *Gramática do português culto falado no Brasil*. Vol. II - Classes de Palavras e Processos de construção. Campinas, Editora da Unicamp, (pp. 623-808).
- Jackendoff, R. (1991) *Semantic structures*. Cambridge, The MIT Press
- Miller, G. A.; Fellbaum, C. (1991) Semantic networks of English. In *Cognition*. Amsterdam, v. 41, (pp. 197-229).
- Ruppenhofer, J.; Ellsworth, M.; Petruck, M.; Johnson, C.; Scheffczyk, J. (2010) *FrameNet II: Extended Theory and Practice*.
- Saint-Dizier, P. (2005). PrepNet: A framework for describing prepositions: Preliminary investigation results. In *Proceedings of the Sixth International Workshop on Computational Semantics (IWCS'05)* (pp. 145-157).
- Saint-Dizier, P. (Ed.). (2006). *Syntax and semantics of prepositions* (Vol. 29). Springer Science & Business Media.
- Taylor, J. R. (1995) Linguistic Categorization. In *Prototypes in Linguistic Theory*. Oxford: Clarendon Press.