Adding Semantic to solve 'PP attachment' in Spanish

Cuevas-Rasgado A.D.[1] , Guzmán-Arenas A. [2] , Juárez-Gambino J.O. [3], Lake-Moctezuma F.L.[3] , Padró-Cirera L.[4]

[1] Universidad Autónoma del Estado de México, Computer Engineering, Estado de México, México

[2] Centro de Investigación en Computación Instituto Politécnico Nacional, Data Science Laboratory, Mexico City, Mexico

[3] Escuela Superior de Cómputo Instituto Politécnico Nacional, Computer Systems Engineering, Mexico City, Mexico

[4] Universitat Politecnica de Catalunya, Computer Science, TALP Research Center, Barcelona, Spain

adcuevasr@uaemex.mx, aguzman@ieee.org, omarjg82@gmail.com, lakemoctezuma@gmail.com, padro@cs.upc.edu

Abstract. This paper describes a semantic module called Disambiguator of prepositions (DISPRE) in English, DESambiguador de PREposiciones in Spanish (hereinafter, DESPRE), that solves the prepositional phrase attachment for Spanish, and complements the syntactic analysis (based on tags and syntactic rules) of project FreeLing. Syntactic analysis is not enough to correctly disambiguate prepositions and attach each prepositional phrase to the appropriate head. This is because, depending on the context, the same preposition conveys different meanings. DESPRE analyzes the preposition and its surroundings, assigning it to its correct sense; once the preposition sense is disambiguated, parsing rules can easily determine which is the right head, thus building the correct dependency tree. In this manner, with the help of the preposition semantics, the sentence is disambiguated, and the prepositional phrase is correctly attached. DESPRE uses an ontology with semantic frames (for each prepositional sense) to disambiguate and solve prepositional phrase attachment. New rules have been added to the FreeLing dependency parser to handle the different preposition senses. Thus, DESPRE, the modified version of FreeLing, uses syntactic and semantic analysis. Several examples show the effectiveness of DESPRE, which was able to attach correctly prepositional phrases that the syntactic analysis of Freeling alone could not do.

Keywords. PP attachment, semantic frame, ontology, natural language processing, syntactic rules.

1 Introduction

Incorrect attachment of Prepositional Phrases (PP attachment) often constitutes the largest single source of errors in current parsing systems [1]. What makes PP attachment particularly difficult is that the ambiguities often cannot be solved using only structural preferences. For instance, consider the sentence *Veo un gato con un telescopio* {*I see a cat with a telescope*}. To whom is the telescope attached, to me or to the cat? It seems that syntactic analysis will not solve the problem completely, and that use of semantic knowledge is needed. For example, the sentence: *El ladrón vio a la policía con un arma* {*The robber saw the police with a gun*} is easier for the computer to disambiguate than *El ladrón vio a la policía con binoculares* {*The robber saw the police with binoculars*}*,* once the meaning of some words is considered.

To provide the computer with better ways to analyze and understand natural language

sentences is the motivation of this article. Syntactic analysis alone will not completely solve the PP attachment problem is the hypothesis of this paper, the sense of the preposition and the meaning of the nouns involved are needed. Thus, further work is required in adding semantic information to the analyzers. DESPRE uses the prepositional phrase semantics to determine the best place of attachment. It does so by adding to FreeLing [2] the semantic information through an ontology, represented as a network where the nodes are the concepts (mainly nouns) and the edges (links) are relations (mainly verbs) between two nodes.

The methodology employed (Section 3) is to use syntactic and semantic analysis where they work best. Thus, Freeling's syntactical analysis first tags the words in a sentence with Expert Advisory Group on Language Engineering Standards (EAGLES) labels (Figure 3). After the preposition is found, the semantic analysis of DESPRE uses knowledge stored in an ontology, and the context surrounding the preposition, to find its correct sense or meaning; that is, to disambiguate it. Then, the preposition is replaced in the sentence by a "sense of the preposition" (conveying the correct sense), and this updated sentence is given back to Freeling to create a correct dependency tree.

This article shows how DESPRE disambiguates 17 Spanish prepositions in their different semantic categories or senses, with senses given by the dictionary of The Real Academia Española RAE dictionary (hereinafter, RAE[1] dictionary). They are: **a** {to, towards}, **ante** {before}, **como** {as, like}, **con** {with}, **contra** {against}, **de** {of, from}, **desde** {since, from}, **durante** {during}, **en** {in}, **entre** {between}, **hacia** {towards}, **hasta** {until}, **para** {for, to}, **por** {by}, **según** {according to}, **sobre** {on, about, above}, **tras** {behind, after}. More importantly, this work shows how the semantic method employed can be used to analyze all of them once suitable frames (Figure 1) are added to the ontology.

The main contributions of this work are:

1. Adding semantics to syntactic analysis is effective for better understanding of sentences involving PP attachment.
2. How to use an ontology to achieve (1), together with a syntactic tool (Freeling). An example in Figure 2.
3. The method does not require annotated corpora, nor an extensive set of data to train a classifier, or to extract statistics.

The structure of the article is as follows. Section 1 presents the theoretical framework; Section 2 contains relevant state-of-the-art papers, and our approach is given in Section 3; DESPRE's algorithms appear in Section 4. Examples and results follow in Section 5, while Section 6 contains the analysis of a complete corpus. Finally, conclusions are found in Section 7.

Cooking pot:
Round pot of metal or clay, which commonly forms a belly, with a wide neck and mouth and with one or two handles, which is used to cook food, heat water, etc.
Shape: round, oval, cylindrical.
Size: small, medium, large, extra big.
Use: to cook, to store food, to roast, to heat liquids.
Contents: liquids, mole, soups, stews like: rice, beans, broad beans, pasta, barbecue, etc.
Place where it is used: kitchen, restaurants, picnic, dining room.
People that use it: chefs, housewives, diners, sellers of cooking utensils.
Place where it comes from: towns, cities, craft markets.

Fig. 1. Frame that describes a cooking pot. Adapted from [3]. The definition was obtained from https://rae.es and https.wordreference.com.

1.1 Frames

Frames were introduced by [3], who defines them as "data structures for representing stereotyped situations, like being in a certain kind of living room, an event or object like a cooking pot. Attached to each frame are several kinds of information". This work uses frames to add

[1] https://www.rae.es/

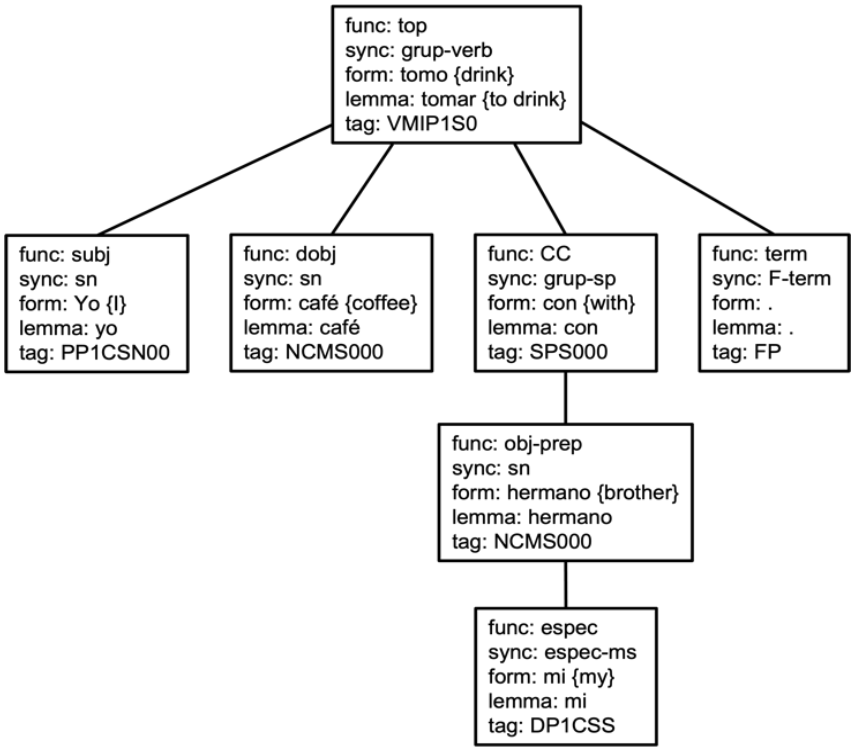

Fig. 2. Dependency tree that FreeLing produces for example d): *Yo tomo café con mi hermano* {*I drink coffee with my brother*}. The preposition *con* {*with*} correctly links *mi hermano* {*my brother*} with the verb *tomar* {*drink*}.

semantic knowledge to DESPRE. Briefly, a collection of suitably related frames forms an ontology. Examples of (parts of) ontologies appear in Figures 4 and 5. For example, the frame for a cooking pot shown in Figure 1 is used to identify the type of relationship between the words of a sentence.

This paper is based on [4], which uses frames as explained below.

Semantic frameworks (Frames) are useful to find the meaning of each word in a sentence, without ambiguities, and ensuring that its syntax and semantics are correctly defined and identified (Figure 1).

They help to identify situations, or the meaning of sentences based on the prepositions used.

1.2 Building the frame for a sense of a preposition

The RAE dictionary gives 26 senses for preposition *de* {of, from}. As an example, let us build the frame for sense 4 (called d4 in Table 1), the RAE dictionary defines this sense as:

d4 Denota la materia de que está hecho algo. *El vaso de plata*. *El vestido de seda*. {Substance; made of}

Sense d4 denotes the matter of which something is made. It is used in phrases such as: *El vaso de plata* {*The silver glass*} or *olla de barro* {*clay pot*}. It describes the material or substance (*clay*) of which certain things (pot) are made of. A frame is wanted to represent this meaning. The frame could be as simple as pot made of clay, but it is necessary to generalize as much as possible this meaning, not only for pots, glasses, or clay. WordNet (WN)[2] comes in handy for this. Pot generalizes (a superset, a hypernym) to: *cooking utensil, *kitchen utensil, *utensil, *implement, *instrumentation, *artifact, *object, *physical entity (Figure 5 shows this part of the ontology), while clay generalizes to: soil *earth, *material, *substance. It is necessary to stop the generalizations so that the sense d4 is expressed as generally as possible, but not more. For instance, substance generalizes to physical entities, but it is clear those *physical entities* are made of substances but not those physical entities are *made of physical entities*. Thus, the frame for sense d4 is *physical entity made of substance*.

[2] https://wordnet.princeton.edu/

When disambiguating *olla de barro*, {*pot of clay*} DESPRE locates *olla* {*pot*} as a physical entity (using ontology shown in Figure 5), and *clay* {*barro*} as a substance, so that the sentence disambiguates to: *olla hecho de* {*made of*} *barro.* The net effect is that *de* becomes *hecho de* {made of}, which is the right meaning in this case. The sense of the preposition *de* {*of*} for sense d4 is *hecho de* {*made of*}*.* These senses of the preposition need to be given to Freeling in a set of adjusted rules (box in the right of Figure 3), so that it can construct a correct dependency tree.

Another sense for *de* is given by the RAE dictionary as sense 5, which in Table 1 is:

d5 Úsase para señalar lo contenido en algo {Contained in something}.

Sense 5 of *de* is used to indicate what is contained in something. For example: *un vaso de agua* {a glass of water}, *olla de arroz* {*rice pot*}. In this case, pot generalizes to vessel with its ancestors: *container, *instrumentality, *artifact, *object, *physical entity, but stops the generalization in this case to container. Rice generalizes to *grain, *foodstuff, *food, *substance, *physical entity, but the generalization stops in food. Therefore, the sense d5 is expressed by frame container: *que contiene* {*that contains*} food. Sense d5 replaces *de* by the "correct preposition" *que contiene.* Then *de* becomes *hecho de* {*made of*} or *que contiene* {*that contains*}, depending on context. Figure 4 shows these two senses of *de.* In fact, RAE dictionary says that *de* has 26 senses, of which DESPRE now disambiguates 13: those shown in Table 1.

1.3 Storing frames in an ontology

[5] explains in detail how an ontology is formed where the Ontology Merging Notation is explained. Briefly, an ontology can be seen as a network where links or edges relate concepts or nodes. Figure 3 show the use of the ontology in which a noun represents a concept that is a node in the ontology, while verbs or actions are edges that link two nodes. Figures 4 and 5 show parts of an ontology. Since each sense of a preposition is stored as a frame in the ontology, which now contains 50 frames: those corresponding to the 50 disambiguated senses. See Table 1.

2 Related Work

Most previous work in disambiguation is for the English language, mainly for nouns and verbs. Regarding prepositions, Yan and Nguyen [6] do not attempt to disambiguate PP attachments, instead, they aim to get a better understanding of the nature of PP attachments by investigating the internal and external factors to the head word sequence in the verb phrase: V-N1-P-N2. They use the PP attachment corpus in English that contains 28,000 PP attachment decisions. Sometimes the lexical elements may determine the interpretation of a given sentence but at other times, they fail. Their proposal uses five hypotheses based on theoretical considerations and initial empirical observations that concern aspects of the elements V-N1-P-N2 (considering external and internal factors). They use the context where the sequence appears. Idiomatic expressions and sentences with two or more prepositions are not considered**.** This work is relevant, but it emphasizes a different perspective from DESPRE, since it uses hypotheses that must be true.

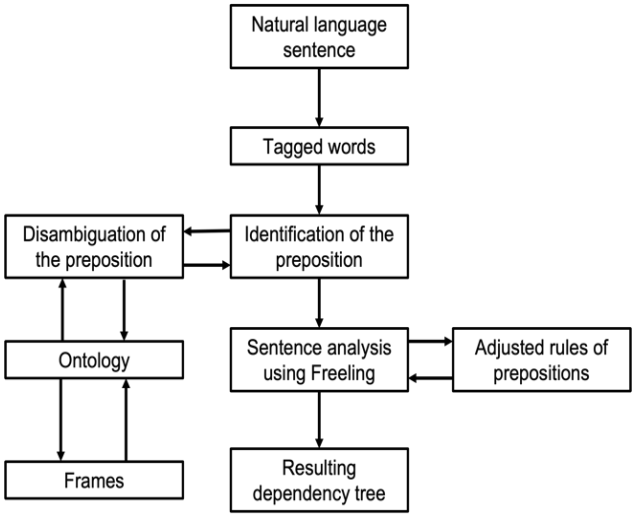

Fig. 3. Diagram of DESPRE. A Spanish sentence enters at the top. It is tagged. The preposition is found. Going to the left of the diagram, the preposition is disambiguated with the help of an ontology, and a suitable replacement for the preposition is returned to FreeLing. The altered sentence is given back to FreeLing, which now produces a correct dependency tree, after considering the new rules (rectangle to the right) added to parse the replacement.

2.1 FreeLing

[FreeLing is a library with a set of Natural Language Processing (NLP) applications, whose purpose is to facilitate the integration of language analysis tools in high-level applications [2]. The online version requests a data entry in natural language text, to which the user chooses the analysis tool he/she wishes to test labeling, morphological analysis, deep analysis, dependency tree, etc. Later, the result of the tool applied to the text is presented.

FreeLing analyzes several languages. The Spanish language has been chosen because it is the most complete and robust module with dictionaries and preprocessing.

The modules that the Spanish language contains are: Lexicographic analyzer, sentence separator, number detection, morphological dictionary, rule applicator, multi-word detection, basic detection of named entities, detection of BIO named entities, quantity detection, labeling PoS, phonetic encoding, WN sense annotation, UKB sense disambiguation, shallow analysis, deep dependency analysis, preference resolution [2].

2.1.1 Interaction of FreeLing with DESPRE

The objective of DESPRE in FreeLing is to add a semantic analysis over the morpho-syntactic analysis prior to the construction of the dependency tree. Specifically, DESPRE identifies if there is any preposition between the elements of the sentence, analyzes if there is a semantic relationship between the nouns that are immediately to its left and right; the semantic frameworks are used for this. These frames will look for a semantic relationship between preposition and nearest nouns. If so, FreeLing should be told how to join the nodes that surround the preposition in the dependency tree.

DESPRE receives a text in natural language to analyze after the module FreeLing's Morpho is accessed. This module contains: number detection, date detection, multi-word detection, dictionary search, number recognition, dates, quantities, punctuation marks, proper noun recognition, word tagger, and Named Entity recognizer. This recognition is based on the Conference on Computational Natural Language Learning (CoNLL) corpus 2002 [7].

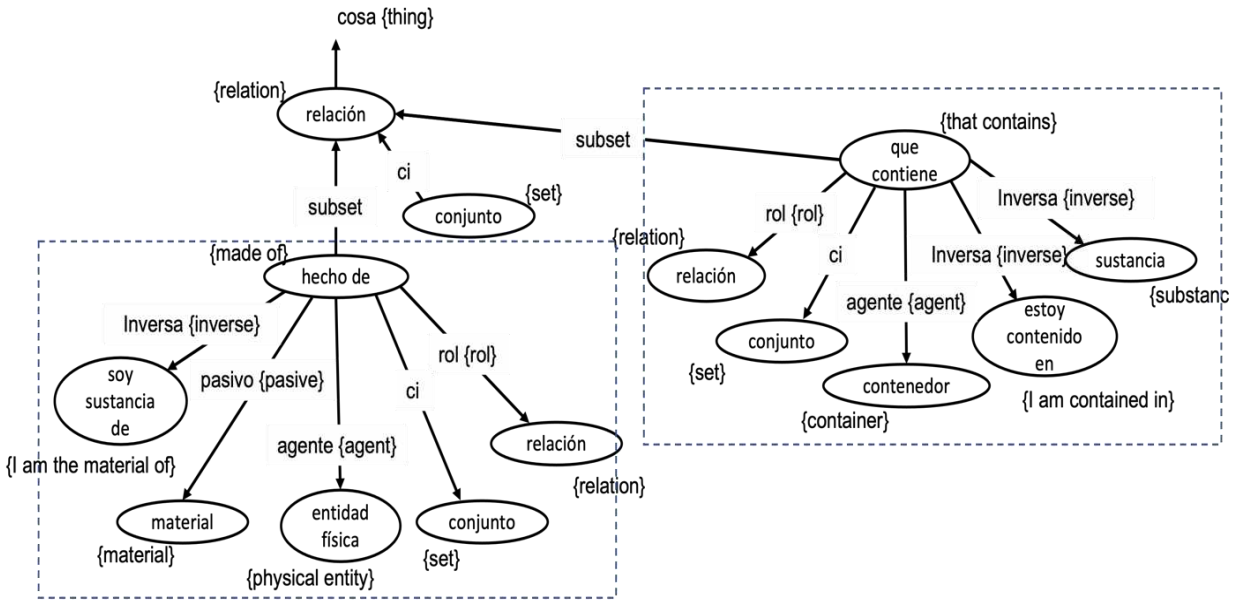

Fig. 4. This part of the ontology shows two meanings of the preposition *de*.

After Morpho labels all the words in the text, DESPRE applies the Prepositions Identification module. Then, the Prepositions Disambiguation module looks for the meaning of the preposition in the ontology, locates the appropriate frames, and replaces the prepositions with the corresponding sense. The tagged text (with the prepositions resolved) is sent to the FreeLing dependency analysis, and the analysis is realized, employing 50 additional new rules. These new rules were adjusted to the senses of the prepositions.
Then, the constructor of the dependency tree is invoked and finally the dependency tree is presented. The graphic description is shown in Figure 3.

2.2 Disambiguation of prepositions in Spanish

[8] describes an algorithm and its implementation to disambiguate prepositions in Spanish phrases; some of them are to {*a*}, before {*ante*}, under {*bajo*}, for {*para*}. They use stand-alone frames, not forming part of an ontology. Their algorithm tags (with EAGLES labels) each word of a sentence. Then, it looks for a preposition and the word before and after it, much as is done now. After that, it looks for the frames of these three words. If at least one 'hypernym' matches, their algorithm takes the synonymy of the preposition, and it uses this sense as the correct one. The difference with DESPRE is that [8] only uses the 'hypernym' of the frame. Since the frames are part of an ontology (see Figure 4), not only the closest 'hypernym can be used', but also any predecessor, until a match is found (refer to Figure 4). Another difference: DESPRE ignores the adjectives before and after each word. For instance, for the sentence: *olla verde bonita de hermoso barro* {nice green pot of beautiful clay} our algorithm reduces it to *olla de barro* {pot of clay} and disambiguates that into *olla hecha de barro* {pot made of clay}, thus finding the correct sense for preposition *de* {of}.

2.3 'PP attachment'

When a parser builds a syntactic tree of a sentence, it faces the problem to decide where to attach a phrase to a list of possible places. An example of this kind of ambiguity is the PP attachment. This kind of ambiguity consists in deciding if the attachment is to a verb or a noun, or to which noun. The PP attachment is considered one of the most difficult constructions for parsers, causing many of their parsing errors [9].

Researchers have proposed several methods to solve 'PP attachment', and one way to classify their proposals is by identifying the previous knowledge they use.

The corpus-based approach uses annotated examples of sentences with PP attachment ambiguity solved, and these examples are used to train a supervised machine-learning algorithm that learns from annotated sense examples. [10] use semantic classes to improve the parsing and 'PP attachment' tasks in English sentences. The authors preprocess the input by substituting some words with their corresponding synsets of WN [11], for example, *knife* is replaced with *tool*, and then they use three different disambiguation methods. The first method uses the manually annotated senses in a gold standard corpus created from merging the Penn Treebank [12] corpora. The second method disambiguates by taking the first sense data from WN. The third uses an unsupervised system to predict automatically the most frequent sense of the words. The results showed for the third method an error reduction rate of 20.5 % in the 'PP attachment'.

In [13], a classifier was trained with different features to decide whether the preposition is attached to a verb or a noun. It works with Spanish sentences. The authors extracted 4,764 examples of 'PP attachment' ambiguities from the AnCora corpus [14] and used 90% for training and 10 % for testing. The EsTxala parser in the FreeLing Environment, which obtained 61% of accuracy, was taken as a baseline. The classifier was the best with 87.84% of accuracy, improving by 26% the results of the baseline.

Unlike these works, DESPRE does not need previously annotated examples to disambiguate the prepositions. This is a very important difference because the effort associated with manual tagging could be very high.

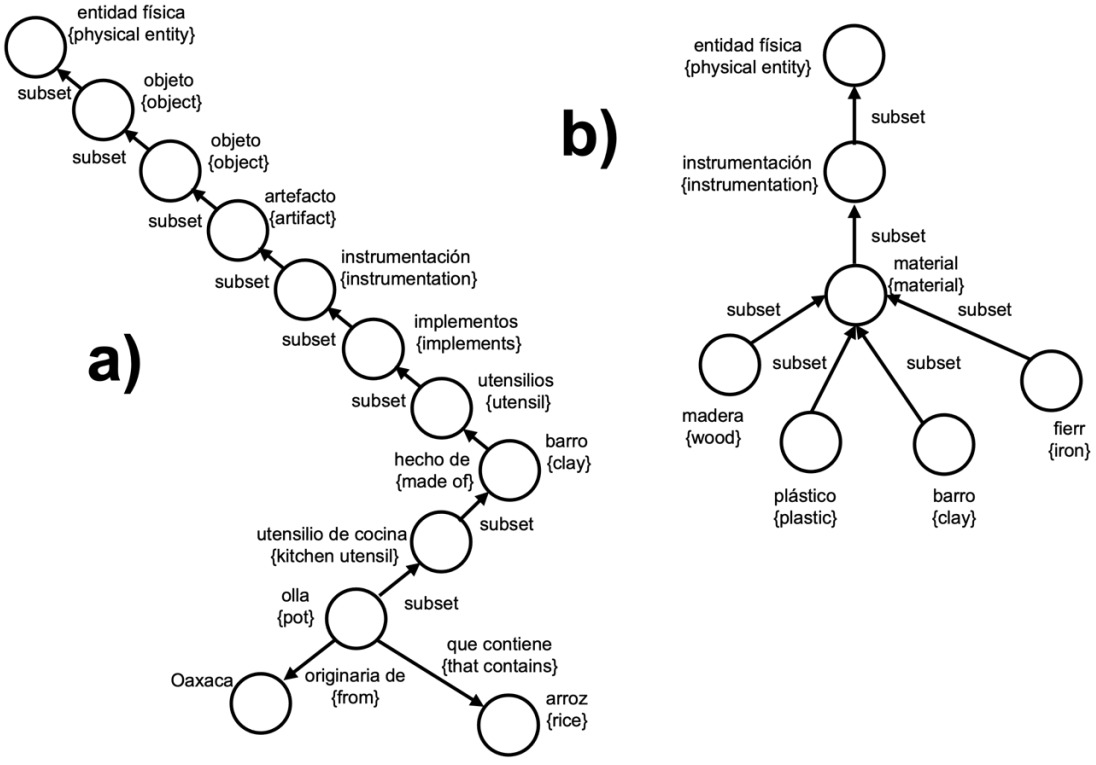

Fig. 5. Another part of the ontology. (a) *entidad física* {*physical entity*} is a predecessor of *olla* {pot}. (b) *material* {*material*} is a predecessor of *barro* {*clay*}.

In [15] the authors consider that prepositions may have not only two head candidates (a verb or a noun) but a set of them. For this purpose, they created a dataset for the German language which considers multiple head candidates. A topological field model of that language was used to find regularities in word order across different clause types in German**,** and this information helped the algorithm to select the correct head of prepositions. DESPRE uses only one head at a time.

The knowledge-based approach uses linguistic resources but does not heed previously annotated examples. Lexical information gathered from large French corpora was used in [16] to resolve 'PP attachment' ambiguities using two methods. In the first method, the author collected around 45,000,000 pairs of verb + preposition and noun + preposition from the Giga-word corpus according to [17] and calculated the Mutual Information (MI). The MI provided a measure of association strength and helped to disambiguate the 'PP attachment'. The second method created a vector-space representation of the context's words, where the Distributional Semantic Model was a frequency count matrix with a window of size 3 for the context's words. The cosine is used as a comparison vector, the higher the similarity between vectors, the higher the possibility for a particular attachment. The best result, using the second method, showed an accuracy of 76.9%.

[18] used word vector representation to build a compositional neural network architecture. They tested it on some English and Arabic datasets. The word vectors were initially obtained from raw text but in following steps, these vectors were enriched by adding information from WN, VerbNet and the words part-of-speech. All this information trained the neural network, obtaining 88.7% accuracy.

The above approaches –and most of the state-of-the-art in 'PP attachment'– share the supervised machine learning method. A fresh approach is that of [19], like DESPRE, uses frames to disambiguate a few prepositions in the Persian language, obtaining a high precision (99.16%) of its system.

[20] says that morph-syntactic information is often insufficient to resolve the ambiguity problem, additional semantic information or even world knowledge is needed. A PP attachment disambiguation for German system based on biaffine attention and contextual word embeddings is presented. This paper shows that modeling all head-dependent pairs jointly (as done in full parsing) allows the system to make more effective use of the training data and is thus superior to modeling PP attachment as a separate task. During disambiguation process an Artificial Neural Net (ANN) was used. This ANN was training on the

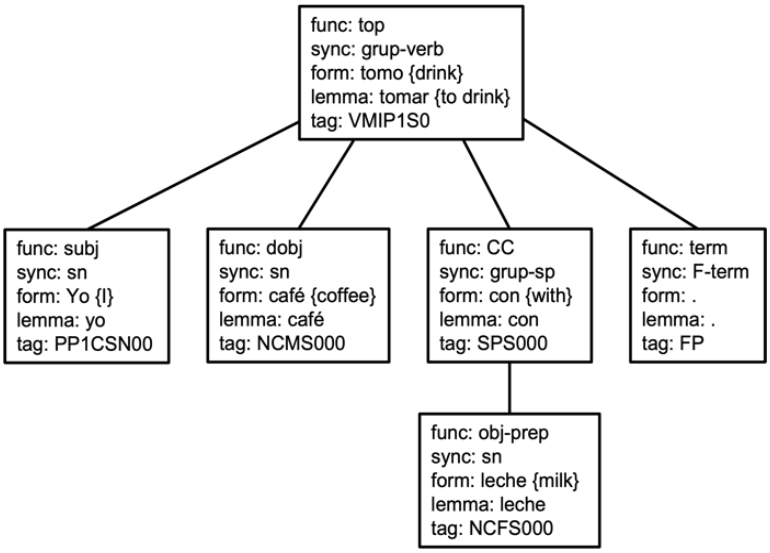

Fig. 6. FreeLing's default output. The preposition *con* {with} relates *leche* {milk} to *tomo* {to drink}, the highest node, which is incorrect.

German dataset from the CoNLL corpus 2009 Share Task. They keep ambiguous and unambiguous triples and calculate five association scores. With the same hyper parameters as PP-REP, their system with auxiliary distributions achieves an accuracy of 86.8%.

In [21] a novel methodology for preposition sense disambiguation (PSD) in English is proposed, in which does not use any linguistic tools. It uses a supervised setting, where the machine learning model is presented with sentences wherein prepositions have been annotated with 'senses'. These senses are IDs called 'The Prepositions Project (TPP)'. A Multi-Layer Perceptron is used to classify IDs into the correct sense. The dataset used is from SemEval-2007 Task-6. The methodology gives an accuracy of 86.85%.

In [22] they implemented a module PATCH-COMM, for various conventional parsers, to make attachment decisions. Where the commonsense KB infers "out-of-knowledge-base" assertions in a similar manner to the way some NLP systems handle out-of-vocabulary words. The results suggest that the commonsense knowledge approach can provide the best of both worlds, integrating rule-based and statistical techniques. Also, PATCH-COMM makes the knowledge acquirable by using Machine Learning mechanisms, even if they may be somewhat opaque. Thus, PATCH-COMM focuses on a particular knowledge base and a particular ANN based architecture for extending the knowledge contained in that knowledge base. His results have been evaluated with Belinkov et. al.'s English dataset (Belinkoc et. Al., 2014 [23]). Only the preposition 'with' was evaluated, having the best result from the paper 88.7%.

In [24], the authors provide a way to resolve English ambiguities under the locally factored

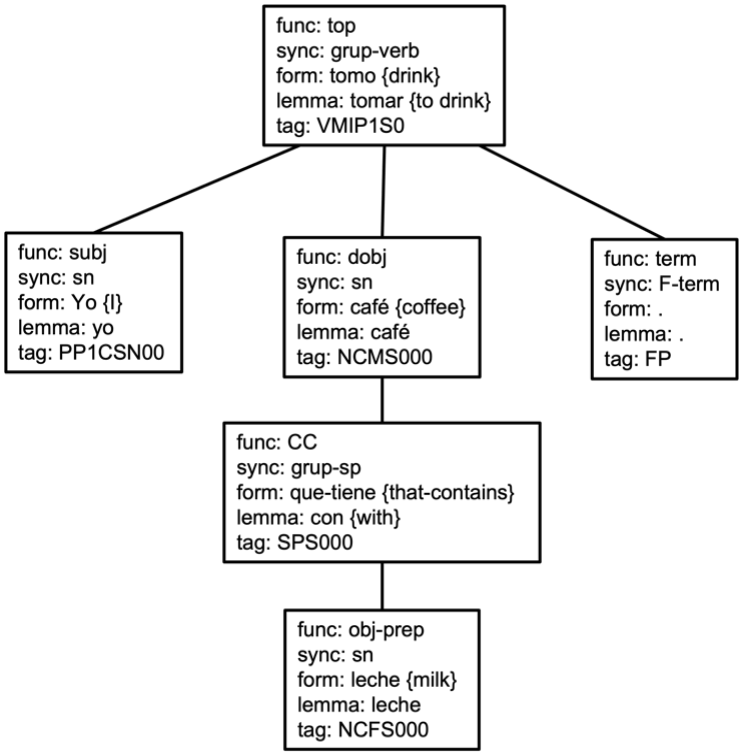

Fig. 7. Correct result, showing that preposition *con* {with} has been replaced by *que tiene* {*that contains*} and linked to *café* {*coffee*}, instead of linking it to *tomo* {*to drink*}.

model by explicitly modeling bi-lexical dependencies. Their proposed model achieves 88.8% in terms of correct labels, which marks the best scores so far.

As is seen in these related works, to solve text analysis problems, it is not enough to use only morpho-syntactic methods. Thus, they are seen complemented with some semantics, through WN, or with classes, tools, specific methods, and algorithms. Thus, FreeLing a semantic module has been chosen. In this way, it can deduce the meaning of words from its presentation in the text and its semantic context to solve their 'PP attachment'. As seen below, the semantic information used by DESPRE improves the 'PP attachment' phase of FreeLing. It could improve, as well, other text analysis tasks, like dependence analysis and reference resolution.

3 Our Approach

3.1 DESPRE

DESPRE focuses on Spanish, since it is an important language, spoken by 580 million people. It is the world's second-most spoken native language after Mandarin Chinese, and the world's fourth-most spoken language overall after English, Mandarin Chinese, and Hindi. DESPRE is centered on only one language, since trying to solve PP attachment in a general way for several languages is much more difficult (for instance, some languages do not have prepositions). In fact, [25] argues that cross-language comparisons of the disambiguation results in previous research is impossible because of the different selection procedures when building the training and test sets.

DESPRE uses a knowledge-based approach, as in [26], to disambiguate the PP

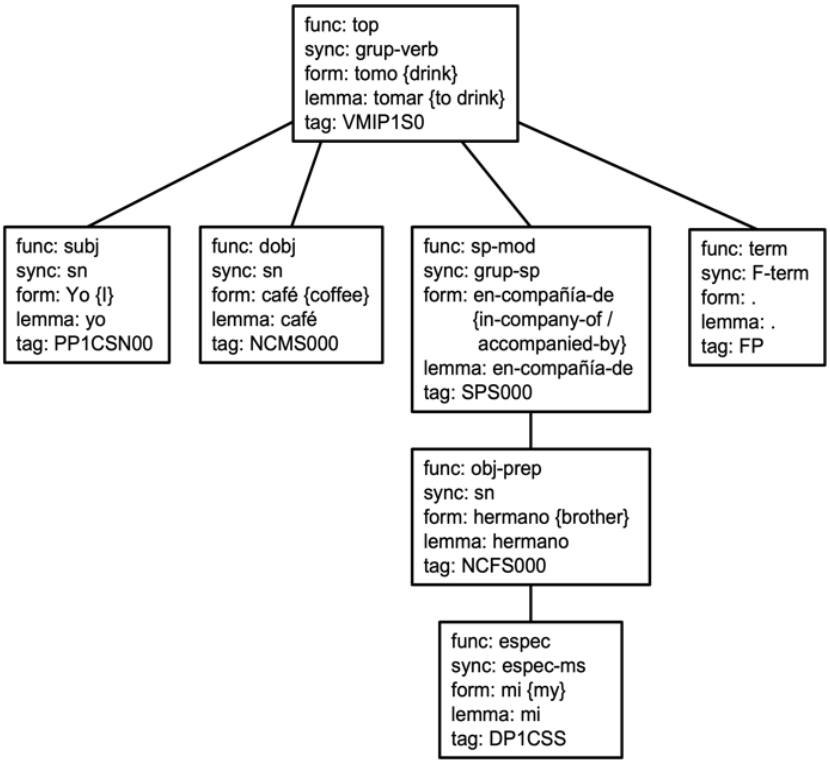

Fig. 8. DESPRE replaces *con* {with} by *en-compañía-de* {accompanied by}, thus linking *mi hermano* {accompanied by my brother} to the verb *tomo* {to drink}, resulting in the correct dependency tree}.

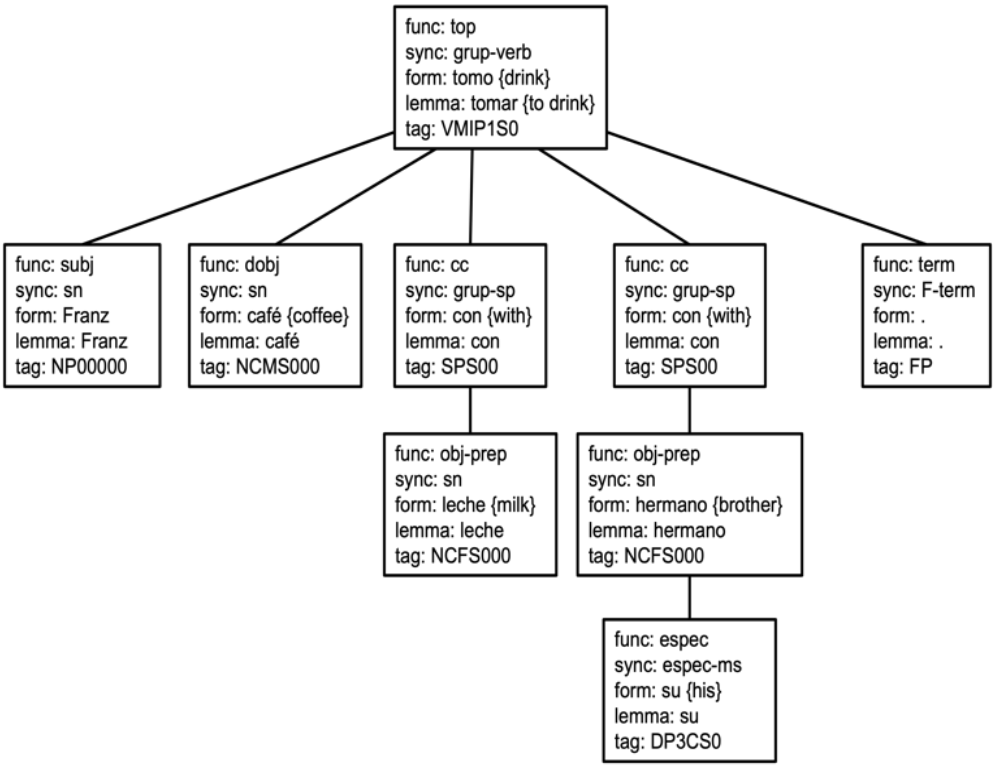

Fig. 9. Normal dependency tree produced by FreeLing for *Franz toma café con leche con su hermano* {*Franz drinks coffee with milk with his brother*}, which is incorrect.

attachment**.** To obtain the sense or meaning of a fragment of a sentence, it employs a populated ontology of prepositions and their different senses**,** instead of using WN or probabilities or statistics.

Freeling is used, since it works very well for Spanish, and it does an excellent job in exploiting its syntax.

DESPRE's strategy is knowledge-based and takes advantage of the frames of the ontology to disambiguate the preposition, so that the parser can use sense-specific rules to properly build the tree. It does not need a training corpus or a huge amount of text to calculate probabilities or mutual information, and it is able to obtain the sense of the preposition from the word's context and the ontology. This facilitates DESPRE's strategy to be re-targeted to other languages where the appropriate resources (frames) are available. In other words, from the frames of the ontology, DESPRE obtains the necessary knowledge to attach the prepositions correctly, while at the same time using to its advantage, as most approaches do, the syntax of Spanish sentences.

3.2 How DESPRE works

The purpose of our work is to classify the 22 Spanish prepositions into their different semantic categories or senses, using the senses given by the RAE dictionary. This paper presents the disambiguation of 17 of them. Nevertheless, this combination of semantic and syntactic methods can be used to analyze all of them once suitable frames are added to the ontology. Their disambiguation permits correct attachment of prepositional phrases, as previously shown in [4].

Our applications added to FreeLing are:

A) Spanish frames that describe the semantic of relations of dependencies.

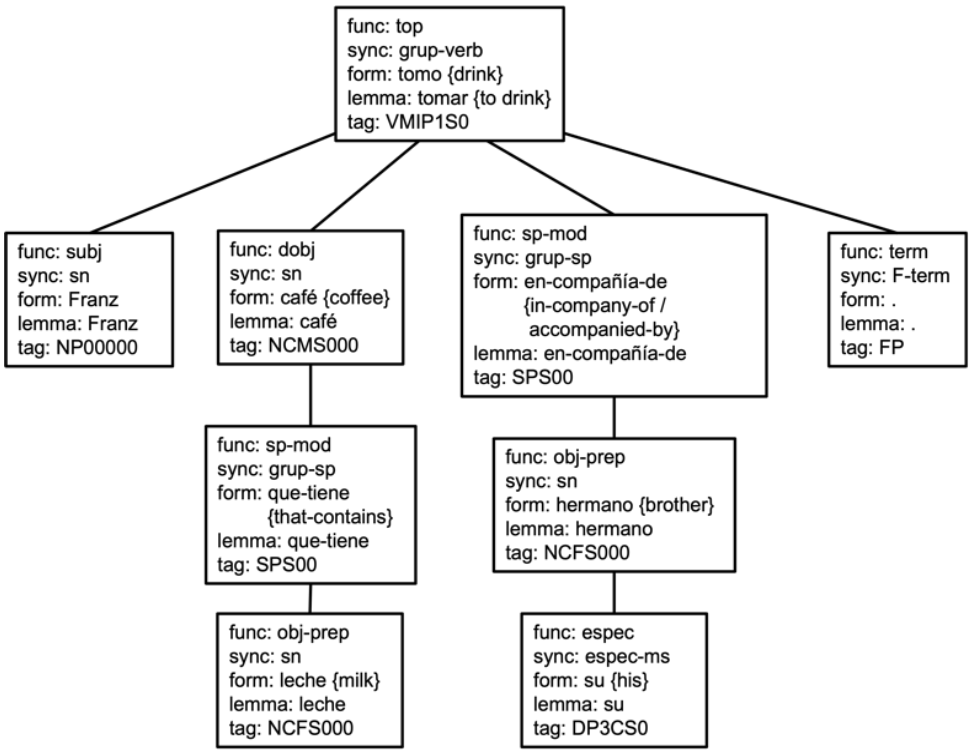

Fig. 10. Correct dependency tree. *Con* {*with*} has been replaced by *que tiene* {*that contains*} when linking *café* {*coffee*} to *leche* {*milk*}. In addition, *con* {with} has been replaced by *en-compañía-de* {*accompanied by*} when linking *toma* {*to drink*} to *en-compañía-de mi hermano* {*accompanied by my brother*}.

B) Source code to connect the semantic analyzer and FreeLing.

C) Source code to connect the semantic analyzer and disambiguate of prepositions.

Our approach is applied after the FreeLing PoS tagger has been run, and before the rule-based dependency parser is executed. Frames are used (they indicate whether there is a semantic relationship between two relevant words) to disambiguate the preposition senses given the context. Then, the tagged sentence with the marked prepositions is fed to the parser, which uses a modified grammar that considers these preposition senses to perform the proper attachment of prepositional phrases. Examples of the use of semantic relationships are:

a) Juan tomó café {*John drank coffee*}, where: tomó {*drank*} represents the relationship between Juan {*John*} and café {*coffee*}. As a handy notation, words joined by hyphens to clarify further the meaning of a preposition are used. For example:
b) *Vaso de agua* {glass of water} typically means: *vaso que tiene agua* {glass that contains water},
c) *mesa de madera* {wood table} means: *mesa hecha de madera* {table made of wood}.
d) *Yo tomo café con mi hermano* {I drink coffee with my brother} typically means: *Yo tomo café en compañía de mi hermano* {I drink coffee accompanied by my brother}.
e) *Yo tomo café con leche* {I drink coffee with milk} means: *Yo tomo café que tiene leche* {I drink coffee that contains milk, mixed with milk}.

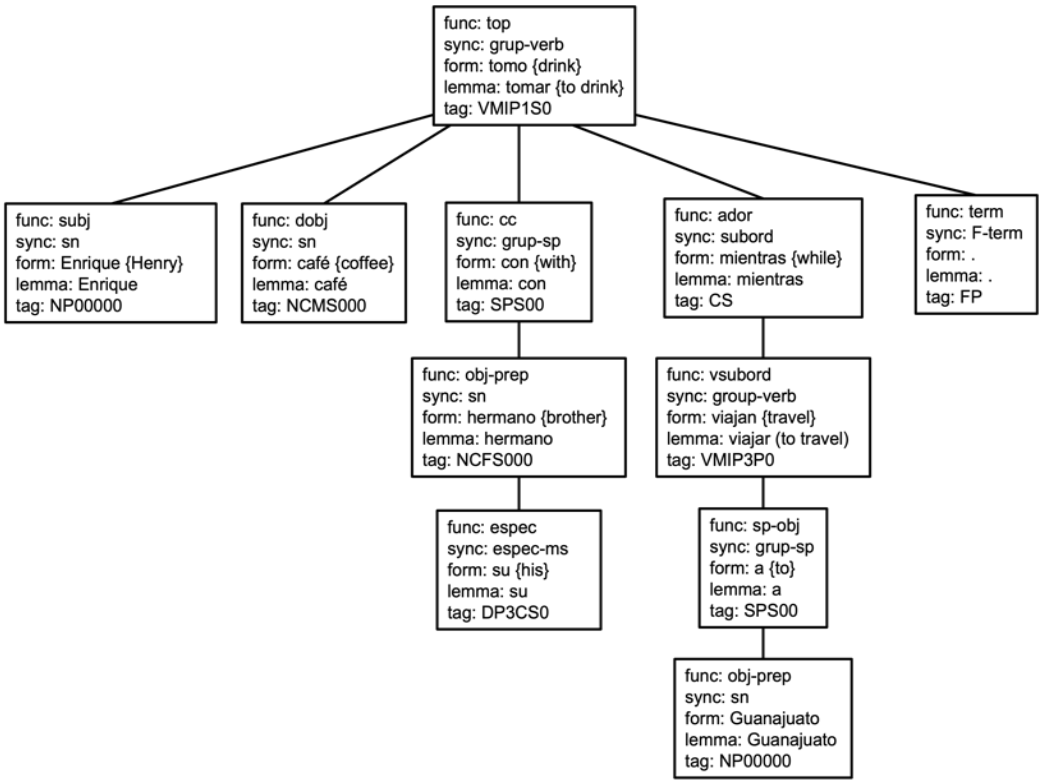

Fig. 11. Usual output of FreeLing for example 4, *Enrique toma café con su hermano mientras viajan a Guanajuato* {Enrique drinks coffee with his brother while they travel to Guanajuato}, showing that *con su hermano* {with his brother} has been linked to *toma* {to drink}, which is incorrect.

Nevertheless, people do not speak with all those details of meaning, because they talk to another person, who readily understands the sentences.

Examples d) and e) have a similar morphological and syntactic structure, since nouns surround the preposition *with* {*con*} in both cases, but the meanings are different. Example d) shows that *café* {coffee} *lo bebo en compañía de* {is drunk accompanied by} *mi hermano* {*my brother*}, while example e) indicates that *café* {*coffee*} *tiene* {*has*} or *está mezclado con* {*is mixed with*}*, que tiene leche* {*milk*}. Figure 2 shows the dependency tree computed by FreeLing for example d**)**, which is correct but lacks the sense of the preposition *con* {*with*}. Figure 8 presents the parsing made by DESPRE, which adds the sense of the preposition *en compañía de* {accompanied by}. However, in example e), milk must be attached to coffee, because the term milk is a modification of coffee. That is to say, drinking coffee *que contiene leche* {that contains milk}. Figure 7 shows the correct parsing (by DESPRE) for example e).

4 Algorithm

The main modules of DESPRE appear in Figure 3. Its main steps are:

1. The ontology is loaded.
2. A Spanish text is entered through the interface, for instance, *La hermosa y brillante olla de legítimo barro negro* {*the beautiful shiny pot of legitimate black clay*}.
3. FreeLing tags the sentence. After that step, it becomes {*La* (DA0FS0) *hermosa* (AQ0FS0) *y* (CC) *brillante* (AQ0CS0) *olla* (NCFS000) *de*

(SPS00) *legítimo* (AQ0MS0) *barro* (NCMS000) *negro* (AQ0MS0). (*Fp*)}

4. DESPRE uses the tags to find prepositions. In the example, it finds *de* (SPS00).
5. For each preposition:

a) The preposition located in the ontology, uses the 3-tuple (left word or active agent, preposition, right word or passive agent). In our example, active agent = *olla* {pot} (NCFS000), preposition = *de* {of} (SPS00), passive agent = *barro* {clay} (NCMS000). The algorithm skips over those words that are not verbs or nouns.

b) For each meaning or sense of the preposition, a frame exists in the ontology.

c) The 3-tuple is compared with each sense. If a match occurs, the correct meaning replaces the preposition in the text. In the example, *olla de barro* {clay pot} is compared with the first of the senses of Figure 4: *hecho de* {*made of*} and *que contiene* {*that contains*}. In this first sense, its agent is *entidad física* {a physical entity}, and its passive agent is {material}. Thus, *olla* {pot} is compared with *entidad física* {*physical entity*} and *barro* {*clay*} is compared with {material}. They do not match directly.

i. If there is no match, the algorithm travels the ontology looking for a predecessor

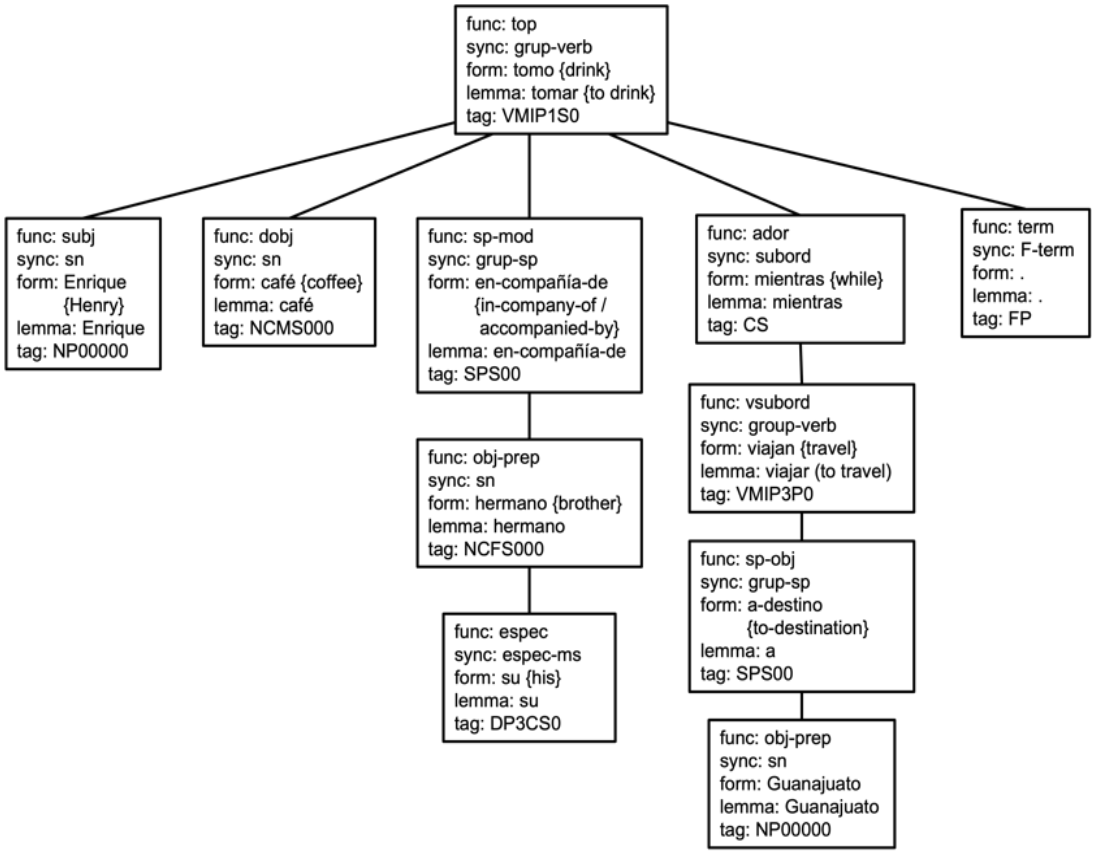

Fig. 12. Correct dependency tree for example 4, *Enrique toma café con su hermano mientras viajan a Guanajuato* {*Enrique drinks coffee with his brother while they travel to Guanajuato*}. *Con* {*with*} has been replaced by *en-compañía-de* {*accompanied by*}, and *a* {*to*} has been replaced by *a-destino* {*to-destination*}.

(hypernym) of *olla* {pot} that matches *entidad física* {*physical entity*}. See Figure 5 (a).

ii. In addition, the algorithm finds that *barro* {*clay*} has a predecessor named {material}. Refer to Figure 5 (b). Therefore, both *entidad física* {*physical entity*} and {material} have matched with the agent and the passive agent, respectively, of the first sense of Figure 4. Therefore, the preposition *de* is replaced by *hecha de* {*made of*} (taken from the first frame of Figure 4), returning the modified phrase: *La hermosa y brillante olla hecha de legítimo barro negro*.

iii. If there is no match, DESPRE analyzes the next sense, returning to b.

d) If all the senses of a given preposition fail, the preposition is not replaced, but it is kept in the sentence. DESPRE has failed for this preposition.

6. The replacement (*hecha de* {*made of*} in our example) is introduced in the sentence. The next preposition is taken, and the algorithm returns to step 5. Once all prepositions are replaced (or left intact, if no match was found in step 5.d), the algorithm goes to step 7.

7. FreeLing performs a syntactic analysis of the modified sentence obtained in step 6. For instance, *Enrique toma café con leche* {*Enrique drinks coffee with milk*} is now sent back to FreeLing as *Enrique toma café que tiene leche* {*Enrique drinks coffee that contains milk*}.

8. FreeLing analyzes the modified sentence and applies the appropriate syntactic rules.

9. FreeLing creates the corresponding dependency tree of the sentence. DESPRE returns the dependency tree, containing the sentence with the prepositions disambiguated.

In summary, FreeLing tags the preposition, then the meaning of this preposition is solved, and the original preposition (for example *con*) is changed to its meaning (for example *que tiene*), preserving the original EAGLES label (SPS00).

Figure 4 shows in the left frame, the node *hecho de* {*made of*} must have as agent a physical entity, such as chair, and its passive element must be a material, such as iron. Thus {table made of

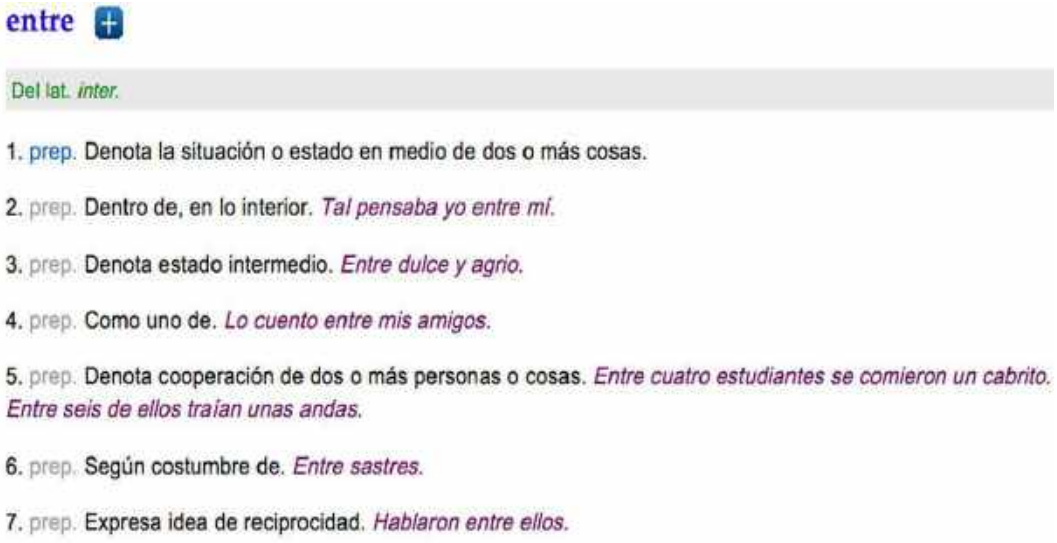

Fig. 13. According to the on-line dictionary of RAE, preposition *entre* has seven senses, shown here. Of these, DESPRE uses the first three for disambiguating the corpus (the rest will be added in a subsequent work). They appear in Table 1, first column, with marks t1, t2 and t3, respectively.

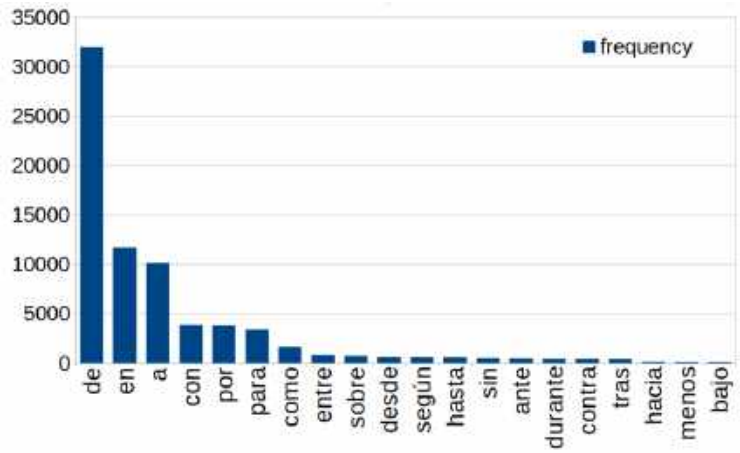

Fig. 14. histogram of occurrence frequency of each preposition in CoNLL 2009 corpus.

wood} is legal, while {polynomial made of iron} or {chair made of plastic} are not, and DESPRE rejects these interpretations. An ontology requires that each node be either a set or an individual. For this reason, the link **ci** (*Conjunto* {set} or *Instancia* {*instance*}) points in both frames to *conjunto* {*set*}, since *hecho de* {*made of*} and *que contiene* {*that contains*} are abstract concepts.

Fig 4 uses a concept called *hecho de* {made of} and Fig 5 a relation with the same name *hecho de* {*made of*}, it refers to the same memory location in the ontology. It could not be represented by other ontology languages.

As said in Section 2.1.1, new FreeLing rules are created for each of these meanings. For instance, if a rule is created for a preposition that is parsed as a verbal group sometimes and as a noun phrase at other times, this rule was created twice, one for each way.

5 Examples

Example 1: *Yo tomo café con leche* {*I drink coffee with milk*}. The usual output of FreeLing (see Figure 6) shows that the preposition *con* {with} links the noun milk {*leche*} with the verb: *tomo* {*I drink*}, which is incorrect (it will be correct to link *popote* {straw} with *tomo* {*I drink*}, as in *Yo tomo café con popote* {*I drink coffee with a straw*}).

The correct meaning is obtained by DESPRE by linking *leche* {*milk*} with *café* {*coffee*}, as Figure 7 indicates.

Example 2: *Yo tomo café con mi hermano* {*I drink coffee with my brother*}. The usual output of FreeLing appears in Figure 2. When FreeLing is used with DESPRE, *con* {*with*} is replaced by *en-compañía-de* {*accompanied by*}, resulting in the correct dependency tree, presented in Figure 8.

Example 3: *Franz toma café con leche con su hermano* {*Franz drinks coffee with milk with his brother*}. Figure 9 displays FreeLing's usual output. Both *con leche* {*with milk*} and *con su hermano* {*with his brother*} have been linked to *toma* {to drink}, which is incorrect. FreeLing's output using the semantic analyzer DESPRE is now shown. The preposition *con {with}* has taken two different meanings: *que-tiene* {*that contains*} and *en-compañía*-de {*in company of, accompanied by*}, as observed in Figure 10.

Example 4: *Enrique toma café con su hermano mientras viajan a Guanajuato* {*Henry drinks coffee with his brother while they travel to Guanajuato*}. Figure 11 shows FreeLing usual output. In Figure 12. FreeLing's output using the semantic analyzer is seen. The preposition *con* {*with*} disambiguates into *en-compañía-de* {*in company of*}, and the preposition *a* {*to*} disambiguates into *a-destino* {*to-destination*}.

For the disambiguation of Spanish prepositions, DESPRE assigns to them their meaning or sense

El gobernante, con [✓ c3] *ganada fama desde* [✓ s1] *que llegó hace 16 meses al* [✓ a1] *poder de* [✓ d11] *explotar al máximo su oratoria y acusado por* [✓ r1] *sus detractores de* [✓ d7] *incontinencia verbal*, *enmudeció desde* [✓ s1] *el momento en* [✓ e1] *el que el Tribunal_Supremo_de_Justicia* (TSJ) *decidió suspender temporalmente los comicios múltiples ante* [✓ n1] *la imposibilidad "técnica" de* [✓ d11] *celebrarlos el 28_de_mayo*. *Chávez se despidió del* [x d13(d6)] *mundanal ruido el pasado jueves con* [✓ c1] *su más breve discurso por* [✓ r8] *televisión, tildado de* [✓ d6] *'institucional' por* [S(r7)] *los observadores, en* [✓ e1] *el que aceptó el aplazamiento de* [✓ d1] *los comicios y valoró la 'pedagógica' medida como* [x a3(c18)] *un triunfo de* [✓ d1] *la democracia venezolana*.

Desde [S(s1)] *entonces entró en* [✓ e1] *silencio absoluto. Nadie sabe cuál es la nueva fecha que_ propone para* [✓ p3] *las votaciones, ni si las quiere juntas o separadas, ni cuándo va a reanudar la campaña.*

Por [S(r4)] *su boca suelen hablar de_vez_en_cuando tanto el ministro de* [✓ d1] *Relaciones_Exteriores,*

Fig 15. The first three paragraphs of the CoNLL corpus 2009. DESPRE has disambiguated all the prepositions of the complete corpus (see https://tinyurl.com/nkexjx2d). Next to each preposition appears its disambiguation, as in *con* [✓ c3] *ganada* or in *del* [x d13(d6)] *mundanal*, indicating correct (✓) or incorrect (x) disambiguation, the correct sense (c3, d6, respectively) and the incorrect sense (d13).

as given by RAE dictionary. Figure 13 gives an example of the seven meanings of preposition *entre* {between}. Since the RAE is regarded as the greatest authority, in terms of use of the Spanish language, just meanings or senses officially recognized by them are considered. The RAE and the *Asociación de Academias de la Lengua Española*, [27] list the following 24 prepositions, two of them no longer used. Next to each preposition appear two numbers, the total senses that the preposition has (as a preposition, since some of them also have meanings as conjunctions, adverbs, etc.), and how many of those senses DESPRE disambiguates now. ***a*** {to} 22/8, ***ante*** {before} 3/1, ***bajo*** {under} 7/0 (only 7 of its 51 senses are prepositions), ***cabe*** (no longer used), ***como*** {as, like} 2/2, ***con*** {with} 7/4, ***contra*** {against} 4/1, ***de*** {of, from} 26/13, ***desde*** {since, from} 3/1, ***durante*** {during} 1/1, ***en*** {in} 8/3, ***entre*** {between} 7/3, ***hacia*** {towards} 2/1, ***hasta*** {until} 3/1, ***mediante*** {through} 1/0, ***para*** {for, to} 10/2, ***por*** {by} 26/5, ***según*** {according to} 3/1, ***sin*** {without} 3/0, ***so*** (no longer used), ***sobre*** {on} 13/2, ***tras*** {behind} 4/1, ***versus*** {versus} 1/0, ***vía*** {via, by} 1/0 (only one of its 16 senses is a preposition).

The RAE dictionary has 158 meanings or sen**s**es of the 22 used prepositions. The current version of DESPRE disambiguates (partially) the 16 prepositions shown in Table 1, into 50 different meanings.

6 Examples using CoNLL corpus 2009

After finishing building the frames of most of DESPRE prepositions, it was decided to test DESPRE on the CoNLL corpus 2009 [7], in Spanish. This corpus was selected because (1) it is in Spanish, and (2) it is easily converted to flat text, the input of DESPRE. Other Spanish corpora exist, but the CoNLL corpus 2019 it is the only one with syntax annotations compatible with FreeLing. Using another will require its previous annotation, a hard task to do. The CoNLL corpus is the result of the effort between work teams, which have the purpose of promoting applications of Natural Language Processing (NLP) and collaborated for 5 years. The corpus is evaluated for it training with the values of the gold-standard annotation [28] of HEAD, DEPREL, PRED and APREDs, and it contain**s** the gold-standard lemma (LEMMA), part-of-speech (POS) and morphological features (FEAT). The CoNLL corpus 2009 Spanish consists of 476,988 tokens in 15,959 sentences (average

Table 1. Prepositions disambiguated by DESPRE.Column headings have the following meaning: **Mark** = the mark that appears in each disambiguated preposition in Figure 15; for instance, c3 in [✓ c3]. **Amt** = how many times this meaning appears in the corpus. ✓= how many times this preposition was correctly assigned this meaning. x = how many times this preposition was incorrectly assigned this meaning. S = how many times DESPRE was unable to assign this meaning to this preposition, for the reason explained in Section 6.

Mark	Meaning and example	Amt	✓/x	S
	Preposition **a** {to; towards}			
a1	**Precede al complemento directo cuando éste es de persona determinada o está de algún modo personificado.** *Respeta a los ancianos. Acabamos de vacunar al perro.* {It precedes the direct complement when it is of a certain person or is in some way personified – Show respect for elderly people. We finished vaccinating the dog}	16	11/1	4
a2	**Precede al complemento indirecto.** *Legó su fortuna a los pobres***.** {It precedes the indirect complement}	6	0/3	3
a3	**Introduce un complemento regido por determinados verbos, adjetivos y nombres**. *Empezar a correr. Disponerse a escapar. Parecerse a alguien. Suave al tacto. Propenso a las enfermedades. Sabor a miel. Temor a las alturas*. {It introduces a complement, governed by some verbs, adjectives, and nouns – To begin to run}	4	2/0	2
a8	**Indica el término de un intervalo de lugar o de tiempo que media entre dos cosas.** *De calle a calle. De once a doce del día*. {Interval end}	1		1
a13	**Ante**. *A la vista*. {In front of; at sight}	1	1/0	
a19	**Por. A instancias mías. {**At my request}	1	1/0	
a20	**Según. A mi entender. A lo que parece. A la moda.** {To my understanding. As it appears. According to fashion}	1	1/0	

29.9 tokens/sentence, minimum 1, maximum 145). Regarding prepositional phrases, the corpus contains 74,410 of them, which represents an average of 4.7 PPs/sentence (minimum 0, maximum 26). It is also noticeable that 72% of PP instances are headed by prepositions *de* (43%), *en {in}* (16%), or *a {to}* (13%). Figure 14 shows a distribution of the occurrence frequency for each preposition.

The Corpus v.2009 is dedicated exclusively to syntactic parsing and semantic dependences, all with a formal analysis based on dependences. The complete disambiguation corpus appears in Figure 15. No person had tagged the prepositions. The corpus came with several named entities (multiword expressions, such as *Tribunal Supremo de Justicia* {Supreme Court of Justice}. To evaluate the performance of DESPRE with the corpus, a couple of competent Spanish speakers were asked to tag (using the tags of column 1 of Table 1) each preposition with the corresponding RAE senses (taken from www.rae.es; a small example appears in Figure 13). In (rare) cases of disagreement, a third native Spanish speaker tilts the balance. The tagged corpus will be considered as a Gold Standard. For clarity, the sense of each preposition appears next to it inside brackets [], containing one of three different symbols:

1. A tick symbol (✓) means that DESPRE assigned its correct meaning.

2. A cross symbol (x) means that DESPRE assigned a wrong meaning; in this case, the true meaning appears inside parenthesis.

3. Letter S means that DESPRE analyzed the preposition but did not assign it a meaning, because the frames needed for this meaning are not built in this version of DESPRE (frames were built for just 50 of the 158 different senses of the prepositions, see example 3 of Section 5).

For instance, preposition *desde {from}* in the first sentence of the corpus shows *desde* [✓ s1] *que llegó*, {since he arrived}, indicating that the correct meaning of *desde* was s1 (see column Mark in Table 1). For instance, preposition *de* {*of*} at the beginning of the second sentence indicates *se despidió del* [x d13(d6)] *mundanal ruido*, {[Chávez] said goodbye to the mundane noise} indicating a mistake (x), the correct meaning, d6, and the incorrect meaning d13 that DESPRE assigned to *del* (a contraction of *de el* {*of the*}). Preposition *por* in the second paragraph displays *por* [S(r7)] *los observadores* {by observers}, indicating that DESPRE was not able to assign a meaning (r7), perhaps because no frames existed to handle this case. The first paragraphs of the corpus appear in Figure 15.

Table 1 shows the prepositions to disambiguate and their multiple meanings (taken from the on-line RAE dictionary and shown as a mark in column *Mark*). Column *Amt* shows the number of times a meaning appears in the corpus. Columns ✓ and show the number of times the prepositions were correctly or incorrectly disambiguated, respectively. Column **S** shows the number of times DESPRE was not able to assign a meaning to the prepositions, due to one of the following reasons (for brevity, the table does not specify which reason applies):

a) The sense for the preposition is no longer used (obsolete meaning); or it refers to an idiomatic phrase containing the preposition (as *de tal palo tal astilla*, {*chip off the old block*}), and its meaning can only be deduced from the whole phrase.
b) The preposition does not link two words, for instance, *de pie* {standing up}.
c) The phrase where the preposition appears plays an adverbial role {l*ocuciones adverbiales*}. Therefore, according to the RAE, they function as an adverb, and their meaning is usually not the combination of the normal meaning of its constituents.
d) The frame of a particular sense of the preposition (such as sense 4 of *entre* {*between*}, see Figure 13) has not been built yet, therefore DESPRE cannot use it.

DESPRE contains frames to disambiguate 50 of the 158 meanings (senses), which correspond to the 22 Spanish prepositions used in Spanish. The results of testing the whole corpus (which contains 230 prepositions) appear in: https://tinyurl.com/auca3e4k Table 1 only shows seven lines. Below shows that DESPRE found the correct meaning in 180 of them, failed in 22 of them, and gave no verdict (column S) in 28 of them.

The accuracy of DESPRE is the percentage of correct PP attachments, compared against the prepositions that it has been designed to handle. Nevertheless**,** currently, DESPRE does not handle all Spanish prepositions [see Section 5 and Table 1], and it would unfair to penalize it when it fails to disambiguate *mediante* {through}, since it lacks the frames for that preposition. When this happens, an S marks this situation in the disambiguated corpus

and in Table 1. Dividing the number of prepositions correctly disambiguated by the sum of correctly and incorrectly disambiguated prepositions yields an accuracy of 89%. As already mentioned, the cases marked with S are not considered, but even considering these 28 cases, accuracy is 78%, which is still considered acceptable. According to RAE (*http://corpus.rae.es/frec/1000_formas.txt*), *de* {*of*} is the most frequent word in Spanish. This preposition is also the most frequent in the analyzed corpus with 96 examples that represent 41% of the total prepositions. Its high frequency is expected to generate more errors during disambiguation, but DESPRE only failed five out of 96 times (that is, 5%) which is a low error rate, showing that it is a robust solution.

7 Conclusions

In a nutshell, syntax is not enough to ascertain the correct meaning of sentences, in a way precise enough for a computer to be able to "understand" written Spanish. Adding semantics as a tool provides much better understanding of the meaning of a sentence. This was proven by disambiguating 17 Spanish prepositions. This was done in a rather simple manner: adding semantic module that interacts with the dependency tree created by FreeLing, a syntactic parser. This combination identifies the meaning of prepositions in sentences. These meanings (usually, a preposition have several meanings or senses) are described in an ontology (following the frames of (Minsky, 1975)), as well as the elements that determine that meaning, that is, the agent (the "leading word") and the passive (the "modifier" of the leading word) associated with the preposition. The method employed by DESPRE does not require annotated corpora, nor an extensive set of training data to learn from it, or to extract statistics. This work can be seen as a semantic addition to a syntactic parser such as FreeLing, an addition that is crucial for our work (which requires precise disambiguation of the senses of words in a sentence). An important conclusion is that FreeLing is adaptable and extensible to the needs of people with other interests besides pure linguistic work. Our main contribution is to develop and implement a semantic module that complements the syntactic analysis (based on tags and syntactic rules) of the FreeLing project that is downloaded by 26,657 users, according to nlp.lsi.upc.edu/freeling/ consulted on March 15th. 2022. The results demonstrate its usefulness; DESPRE assigns to the analyzed prepositions their correct meanings in many cases. In the prepositions of the CoNLL corpus 2009, DESPRE achieved an accuracy (percentage of correct answers) of 89%. In an earlier work [4], the number of disambiguated meanings was 37; in the current work, it is 50. For brevity, only a few examples are shown, but building more frames for disambiguation rules can be continued for the remaining prepositions. Our plan is to implement the remaining meanings of each preposition as given by the RAE dictionary. DESPRE continuation is described in ANACONJ [29] in which an algorithm is shown to identify conjunctions in paragraphs in Spanish using semantic frames.

References

1. **Merlo, P. & Ferrer, E. (2006)** The notion of argument in prepositional phrase attachment. *Computational Linguistics* 32 (3), 341–378.
2. **Padró, L. & Stanilovsky, E. (2012).** FreeLing 3.0: Toward wider multilinguality. *Proceedings of the Language Resources and Evaluation Conference "LREC 2012"* Istanbul:2012, 2473–2479.
3. **Minsky, M. (1975),** A framework for Representating Knowledge, *MIT-AI Laboratory Memo 306,* reprinted in The Psychology of Computer Vision, P. Winston, McGrawHill Ed.
4. **Arana-Roa, B., Lake-Moctezuma F. & Vergara E. (2015).** Implementing a module for semantic analysis upon the syntactic analyzer FreeLing. *B. Sc.Thesis Report 2014-A024,* Escuela Superior de Cómputo, IPN. Mexico.
5. **Guzmán-Arenas, A., and Cuevas, A. D. (2010).** Knowledge accumulation through automatic merging of ontologies. Expert Systems with Applications, 37 (3) 91-105
6. **Yan, M. and Nguyen, M.H. (2018),** The spy saw a cop with a telescope: Who has the telescope? –An attempt to understand the basic building blocks of ambiguous PP attachment sequences, *Journal of Language Works, 3(2).*
7. **Hajic, J., Massimiliano C., Richard J., Daisuke K., Martí, M.A. Márquez, L. Meyers, A., Nivre, J., Padó S., Stepánek J., Stranák, P., Surdeanu, M., Xue, N., and Zhang, Y. (2009)**. The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Lenguages, *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL):* Shared Task, pp. 1-18, Association for Computational Linguistics.
8. ***Villanueva, D., Cuevas, A.D. Juárez, O. & Guzmán, A. (2013).*** Using frames to disambiguate prepositions. *Journal Expert Systems with Applications* Vol. 40, ISSN 0957-4174, pp. 598–610.
9. **de Kok, D., Corina D., Jianqiang M. & Erhard H. (2017).** Extracting a PP attachment data set from a German dependency treebank using topological fields. *Proceedings of the International Workshop on Treebanks and Linguistic Theories,* 89–98.
10. ***Agirre, E., Baldwin, T. & Martinez, D. (2008).*** Improving Parsing and PP Attachment Performance with Sense Information. *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies,* 317-325.
11. **Miller, G. (1995).** Building semantic concordances, *AAI Technical Report. SS-95-01. www.aai.org*
12. ***Mitchell, M., Santorini, B. & Marcinkiewicz, A. (1993).*** Building a large annotated corpus of English: the Penn treebank. *Computational Linguistics, 19(2), 313-330.*
13. **Aguilar, N., Alonso, L., Alemany, M.L. & Castellón I. (2010).** Resolving prepositional phrase attachment ambiguities in Spanish with a classifier. *Procesamiento del Lenguaje Natural, [S.l.],* 46, 75-82, ISSN 1989-7553.
14. **Taulé, T., Martí, M.A. & Recasens, M. (2008).** *A*nCora: Multilevel Annotated Corpora for Catalan and Spanish. *In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08).*
15. ***de Kok, D., Jianqiang M., Corina D. & Erhard H. (2017).*** *PP Attachment: Where do WeStand?. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics,* (2), 311-317.
16. **Šuster, S. (2008).** Resolving PP attachment ambiguity in French with distributional methods. *MSc thesis,* University of Groningen, Nederland and University of Lorraine, France.
17. **Graff, D., & Christopher C. (2008).** *English Gigaword [Online].* Available on: https://catalog.ldc.upenn.edu/LDC2003T05.
18. **Belinkov, Y., Tao L., Regina B. & Globerson, A. (2014)**. Exploring Compositional Architectures and Word Vector Representations for Prepositional Phrase Attachment. *Transactions of the Association for Computational Linguistics,* 2, 561-572.
19. **Mozaffari, Z., Giti, T., Mojtaba S.J. & Pakzad, Y. (2018).** *Preposition sense disambiguation in Persian using semantic frames. Journal of Language Researches* (9)1, 99-117.
20. **Do, B. N., & Rehbein, I. (2020)**. *Parsers know best: German PP attachment revisited* (pp. 2049-2061). ACM Digital Library

21. **Pawar, S., Thombre, S., Mittal, A., Ponkiya, G., & Bhattacharyya, P. (2021)**. Tapping BERT for Preposition Sense Disambiguation. *arXiv preprint arXiv:2111.13972*.
22. **Xin, Y., Lieberman, H., & Chin, P. (2021)**. Revisiting the Prepositional-Phrase Attachment Problem Using Explicit Commonsense Knowledge. *arXiv preprint arXiv:2102.00924*.
23. **Belinkov, Y., Lei, T., Barzilay, R., & Globerson, A. (2014)**. Exploring compositional architectures and word vector representations for prepositional phrase attachment. *Transactions of the Association for Computational Linguistics*, *2*, 561-572.
24. **Masashi, Y., Noji, H. & Matsumoto, Y. (2019).** A * CCG Parsing with a Supertag and Dependency Factored Model, *Journal of Natural Language Processing*, 26(1).
25. **Volk, M. (2006).** How bad is the problem of PP attachment?: a comparison of English, German and Swedish. *Proceeding Preposition '06 of the Third ACL-SIGSEM Workshop on Preposition,* 81-88
26. **Shaidah J., (2018).** A Study on NLP Applications and Ambiguity Problems, *Journal of Theoretical and Applied Information Technology,* 96(6).
27. **Espasa, B. (2011).** Nueva gramática básica de la lengua española, ISBN 978-84-670-3471-4, pp. 164.
28. **CoNLL (2009).** Shared Task: Syntactic and Semantic Dependencies in Multiple Languages, last visit July 2021 https://ufal.mff.cuni.cz/conll2009-st/trial-data.html
29. **Cuevas A. (2018).** ANACONJ Analyzer of the conjunction AND in Spanish using syntactic patterns and semantic frames. Pattern Recognition: Recent Advances and Applications, Mathematical Problems in Engineering, HINDAWI, Vol. 2018 pp: 14.

Article received on 25/07/2021; accepted 01/03/2023.
Corresponding author is Alma Delia Cuevas Rasgado.