

Using frames to disambiguate propositions

Diana Villanueva [Escuela Superior de Cómputo (ESCOM), Instituto Politécnico Nacional, (IPN)]

Alma-Delia Cuevas-Rasgado [Instituto Tecnológico de Oaxaca],

Omar Juárez [ESCOM-IPN], and

Adolfo Guzmán-Arenas¹ [Centro de Investigación en Computación, IPN]

Summary- In Natural Language Processing (NLP), disambiguation is the procedure used to solve name conflicts of polysemic concepts (different meanings); in fact, phrase disambiguation is a problem not totally solved in NLP. Several disambiguation types exist; for instance, noun disambiguation (bank is a word with several meanings) or when a word is both used as a noun and as a preposition. Work herein reported describes an algorithm and its implementation to disambiguate prepositions in Spanish phrases; some of them are: a {to²}, ante {before}, bajo {under}, cabe {fits}, desde {from}, contra {against}, en {in}, por {by}, según {according to}, sin {without}, entre {between}, so {under}, con {with}, hacia {towards}, sobre {about}, hasta {until}, de {of}, tras {after}, para {for}. This tool will be used in a text analyzer (being built) that converts a text document to its corresponding ontology representation, with the goal of allowing another program to use the information to answer non trivial questions.³

Key words. I.2.8. Problem Solving, I.2 artificial Intelligence, I.2.7. Natural Language Processing, Text Analysis.

I. INTRODUCTION

Nowadays much information is in text form, be printed documents (books, magazines) or in digital format. To process it and to represent it in some language understood by a computer, a natural language treatment is needed. Computational Linguistic is the Computer Science area that studies this text processing. Some of its areas are:

- Design of syntactic analyzers.
- Design of taggers and lematizers. Diseño de etiquetadores y lematizadores.
- Design of disambiguators.
- Automatic translators.

¹ Corresponding author. Email: a.guzman@acm.org

² Translation to English are in green for explanation purposes; they gives only one of the senses of the Spanish word, which usually have multiple meanings.

³ This work is part of the project “Use of frames to convert text to semantic networks”, being developed by CIC-IPN and ESCOM-IPN under grant 128163 from CONACYT. The project is continuation of the Ph. D. thesis of one of the authors (A. Cuevas), which automatically fuses ontologies (with no user intervention). The project seeks to automate the text analysis and the ontology builder, too; for this reason, cleaning or preprocessing the document is necessary, and the disambiguator is one of the preprocessing tools.

This article focuses on the design of a preposition disambiguator, using a verb lemmatizer that uses frames [Molina, A. 2004] to identify the meaning of verbs and nouns.

Before a computer can “understand” phrases in natural language, it is important that they are free of ambiguities, and that their syntax be correct.

Spanish is a complex language because interpretation of sentences depends on their grammatical structure, as well as in the precision and elaboration of their rules of formation [MORENO, L., et al. 1999]. Accompanying problems are syntactic and semantic ambiguity (polysemy or different meanings for a sentence, noun, verb, proposition, etc.) [MOONEY, RAYMOND J. 2003].

The article is organized as follows: Section 2 describes concepts used in text analyzers; Section 3 reviews syntactic disambiguators, as well as concepts of our study case; Section 4 proposed a disambiguation method for propositions in Spanish texts; it identifies ambiguous propositions and provides the correct meaning, using the containing phrase as context; Section 5 presents results, applications and difficulties.

II. THEORETICAL BACKGROUND

Conceptually, a NLP system divides text analysis in several layers: morphological analysis, syntactic analysis and semantic analysis [MOONEY, RAYMOND J. 2003]:

- **Morphological analysis:** It determines the form, class or grammatical category of each morpheme (words, affixes, parts of speech) forming a phrase; this is also called morphological tagging.
- **Syntactical analysis or parsing:** It determines the grammatical structure of a text with respect to a given (more or less) formal grammar.
- **Semantic analysis:** It assigns meaning to the syntactic structures; that is, it establishes correspondences between the syntactic structures and the semantic meaning (correct sense) of each work [Palomino K., Rosero R., Zapata C. 2007]

The variability of the Spanish grammar and the precision and elaboration of formation rules difficult the analysis tasks [ZAPATA, C., ARANGO, F.], since several types of ambiguities are present:

- **Morphologic ambiguity:** A word in a sentence may represent more than one syntactic role or grammatical category [MOONEY, RAYMOND J. 2003], [ALLEN, J. 1987], [HAUSSER, R. 2001]. For instance: *Ana no quiere a su esposo porque es infiel* {*Anne does not love her husband because is unfaithful*}. Who is unfaithful? Ana or her husband?
- **Syntactic ambiguity:** When a sentence has more than one syntactic representation (given by more than one grammar rule) [MOONEY, RAYMOND J. 2003], [ALLEN, J. 1987], [HAUSSER, R. 2001]. Example. *En el temblor las mujeres y los niños asustados fueron los primeros en salir del edificio* {*In the earthquake, scared women and children were the first to leave the building*}. It is not clear whether both women and children were scared, or only women. Classes of syntactic ambiguity are:
 - o Coordinative ambiguity. When a sentence contains more than one conjunction: linking conjunctions (and, nor, that), disjunctive conjunctions (or, but, at least, even) and others.
 - o Prepositional ambiguity. When a sentence contains a proposition. This article deals with this type of ambiguity. Example: *Voy a la playa con rocas* {*I go to the beach with rocks*}. Who has the rocks? Me, or the beach?

- **Semantic ambiguity:** When a sentence has more than one meaning; it refers to phenomena like homonymy and polysemy, where a word can have different meanings [MOONEY, RAYMOND J. 2003], [ALLEN, J. 1987], [HAUSSER, R. 2001], [SUÁREZ, CUETO A. 2004]. Example: *Le compré flores* {He bought flowers to her}: the indirect complement *le* may refer to the recipient of the flowers, or to the seller. Example (ambiguity of nouns and verbs): *Sal de México*, {get out of Mexico, or Mexican salt}, where *Sal* may be interpreted as a verb {to leave} or as a noun {salt}. Example (ambiguity caused by propositions; this is the problem solved in this paper): *Olla de barro*, *Olla de México*. {clay pot, Mexican pot}. “De” means “made of” in the first sentence, and “originary from” in the second sentence.

In this paper we deal with semantic ambiguity.

II. RELATED PREVIOUS WORK

Different solution strategies are needed, according to the type of ambiguity present in a sentence [GALICIA, HARO S. 2000].

The disambiguator of [Colorado F. 2008] disambiguates Spanish words. It uses the “Simplified Lesk Algorithm [Vasilescu et al. 2004] with substantial modifications and extensions. In contrast to other disambiguation models, it does not carry a previous tagging of the text, nor it is told which words need disambiguation. It is a non-supervised method. The results she obtains show the model’s strength, when compared to the best international criteria.

Different approaches [Colorado F. 2008], [McCarthy 2004] and [R. Mihalcea - T.Pedersen 2005] classify disambiguation in several groups:

- 1) Statistical methods, which require large computational and lexical resources;
- 2) Knowledge based methods [Cuevas 2006] and Artificial Intelligence methods, using ontologies or semantic networks, and often restricted to specific domains;
- 3) Corpus based methods;
- 4) Hybrid and Bootstrapping methods, presenting a mixture of problems from (1) and (2). Details in [Palomino K., Rosero R., Zapata C. 2007].

Methods of noun and verb disambiguation [Vasilescu et al. 2004; Colorado F. 2008; McCarthy 2004; R. Mihalcea - T.Pedersen 2005] can also use supervised and unsupervised training. These works disambiguate nouns or verbs in a sentence, while our work disambiguates prepositions. As introduction, we give a definition for preposition.

Prepositions

Prepositions are words used to relate ideas composing a sentence; thus, by themselves, they lack expressive meaning of their own, but they contribute in a fundamental manner to provide sentence to the whole sentence since they establish a semantic relation. Many times, prepositions are wrongly used and they form part of the vices of a language.

Classification of propositions

There are two types: the **separable propositions**, more frequent, formed by whole words; and **inseparable propositions**, attached to the word they qualify.

This work studies the **separable propositions**, which are: *a* {to}, *ante* {before}, *bajo* {under}, *cabe* {beside}, *desde* {from}, *contra* {against}, *en* {in}, *por* {by}, *según* {as}, *sin* {without}, *entre* {between}, *so* {under}, *con* {with}, *hacia* {towards}, *sobre* {upon}, *hasta* {until}, *de* {of}, *tras* {behind}, *para* {for}.

Prepositions are classified in groups according to the sense they provide to the expressions they qualify, although some prepositions belong to more than one group:

1. **Place propositions:** Assign a place: *a* {to}, *de* {from}, *en* {in}, *entre* {between}, *hacia* {towards}, *por* {by}, *tras* {behind}.
2. **Time propositions:** A success is place on the time line: *a* {to}, *con* {with}, *de* {of}, *desde* {since}, *en* {in}, *para* {before}, *por* {by}, *sobre* {after}.
3. **Causal propositions:** Provide a link between origin and result: *de* {of}, *por* {by}.
4. **Purpose propositions:** Determinan una finalidad u objetivo: *a* {to}, *para* {for}.
5. **Company propositions:** Expresses association of subjects with respect to an action: *con* {with}.
6. **Instrument propositions:** They express the means by which the action is or has been performed: *a* {to}, *con* {with}, *de* {of}, *en* {in}.
7. **Modal propositions:** They indicate the way in which the action is carried out: *a* {to}, *con* {with}, *de* {of}, *en* {in}, *por* {by}.

Inseparable propositions: Those that join a word to skew or modulate its meaning or orientation, as shown in Table 1.

Preposition	Example		preposition	Example	
a	amoral	{amoral}	peri	periférico	{outlying}
ab	absorber	{to absorb}	poli	polifacético	{versalite}
ante	anteponer	{to put before}	pos	pospuesto	{postponed}
anti	binacional	{binacional}	pre	prenatal	{prenatal}
bi	antihigiénico	{unhygienic}	pro	prosecretario	{assistant secretary}
des	desconfiado	{distrustful}	re	recaída	{relapse}
di	difundir	{to spread}	res	restablecido	{restored}
dis	disfuncional	{dysfunctional}	retro	retrógrado	{retrograde}
equi	equivalente	{equivalent}	sin	sinsabores	{troubles}
extra	extraordinario	{extraordinary}	sub	subsecuente	{subsequent}
hiper	hipertrofiado	{hypertrophied}	super	superfino	{extra fine}
in	inmoral	{immoral}	trans	transoceánico	{transoceanic}
inter	interespatial	{interspatial}	ultra	ultramarino	{overseas}
mono	monaural	{monaural}	uni	unilateral	{unilateral}
o	oponerse	{to object}	vi	virrey	{viceroys}
ob	obsecuente	{obsequious}	vice	vicepresidente	{vice president}
per	perseguir	{to pursue}			

Table 1. Inseparable propositions (first and third columns), with examples. Explanations in {} are for clarity.

Few documented works exist regarding algorithms to disambiguate propositions, and very few to disambiguate Spanish propositions. One of the former is [Palomino K., Rosero R., Zapata C. 2007]. It focuses primarily on syntactic disambiguation, of the prepositional and coordinative type; results are interesting; their approach to the problem differs from ours as follows: Their method applies four steps: (1) the syntactic analysis of the sentence, using a module of NLTK [Website 4],

a set of tools for natural language processing written in Python [Web site 5]; (2) the type of syntactic ambiguity in the sentence is determined, coordinative or prepositional; (3) this is the disambiguation step, which depends on the type of ambiguity detected in (2); (4) results are shown. Our work does not take into account the syntactic ambiguity, but it assumes instead that the syntax is correctly detected, and we only deal with semantic ambiguity (that is, the sense or the meaning of the sentence).

Another work [Nübel R. 1996] deals with disambiguation of prepositions for multilingual spoken translation (German, English and Japanese): one person is fluent in German, another in Japanese, and the translator among them gets help from VERBMOVIL (a real-time translator of dialogues; undergoing tests) in English. It uses lexical-semantic information about the structure of the dialogue, to ascertain how much the meaning of prepositions depends on context. The relation to our work is that they use ontologies to map abstract relations (templates) that help to determine the context of the proposition. Each proposition is linked to a template or abstract relation that indicates the possibility of the use of that proposition in the context of the template. The model of the dialogue is being built through a finite state machine.

In [Litkowski K. and Hargraves O.] a method is exposed to disambiguate prepositions, applied to lexical examples. It uses a set of 25,000 lexical instances comprising the 34 most frequent English prepositions. Each instance identifies a proposition (in a complete sentence) to be tagged (EAGLE tags), taken from FrameNet, a corpus based on the Oxford Dictionary, used to identify the coherence of prepositions in sentences. The method is a supervised algorithm rendering 0.693 accuracy.

In our work, the semantics of the words is inside the description of the frames, which help to identify the context surrounding the proposition.

Frames

Frames were perhaps introduced by [Minsky, 1975], who defines them as data structures to represent generic situations, such as the representation of a type of behavior and things expected in a living room, in a restaurant, in a children party. Several kinds of information are found in frames: how to use it; what activities or successes are performed there; what other successes occur after the frame (a subsequent frame). A frame can be considered as a network of nodes and relations. Frames represent scenes, for instance: the frame representing a children party has more semantics than the definition in a dictionary: [*fiesta* -- *Reunión de gente para celebrar algún suceso, o simplemente para divertirse (Diccionario de la Real Academia Española)*]. party: a social gathering; also : the entertainment provided for it. (Webster Dictionary)]. Table 2 elaborates the frame “children party”.

<p>Dress: better than everyday dresses. Gift: Must be wrapped and given to the celebrated child. Celebrating hide-and-see, guessing riddles, jokes. Decoration: balloons, figures made of crepe paper. Cake: candles, wishes, to blow. Songs: Happy Birthday. Place: Party room, home, kindergarten.</p>
--

Table 2. Part of the frame “children party”.

Many concepts may be related to a frame, and they will enrich it semantically; thus, most of the information intertwined in the represented scenario (concept, event, or activity) must be present in

the frame. There is no standard notation about frames; we have proposed one (Table 3). It is formed by:

- 1) **Header**, representing the name of the concept, a character indicating if it is a noun (s, for *sustantivo*), verb (v), etc., and in parenthesis its gender.
- 2) **Body**, containing elements such as: *quees* {what it is; its meaning}, *agenteactivo* {active agent, who initiates or carries out the action}, *agentepasivo* {passive agent, who suffers the change or consequences of the action provoked by the active agent}, *hiperonimia* {hypernym, concept of which the frame is a subset}, *sinónimo* {synonym, other words used for this frame}, *hiponimia* {hyponym, a particularization of this frame, a frame which is a subset of this frame}, *comoseusa* {how it is used; example of the concept in a sentence}, *frasetemat* {thematic phrase, colloquial sentence, where the word is used with a different meaning}, and
- 3) Our notation allows adding semantic elements as needed, for instance the element *lugar*. {place}.

(defmarco fiesta infantil s (genero fem)	{children party} {female gender}
(quees "Reunión de personas en un lugar para divertirse o celebrar un acontecimiento y en la que se suele bailar, comer, etc.")	{what it is}
	{Gathering of persons in a place to have a good time or to celebrate an event and in which it is customary to dance, to eat, etc.} ⁴
(agenteactivo "El que organiza la fiesta", "el festejado")	{active agent}
(agentepasivo "El que acude a la fiesta", "familiares", "amigos")	{The celebrated person}
(hiperonimia "acontecimiento", "suceso", "Reunión")	{passive agent}
(sinonimo "fiesta de niños")	{Who comes to the party, relatives, friends}
(hiponimia "fiesta de cumpleaños", "celebración de bautizo")	{hypernym}
(lugar "sala de fiestas", "casa particular", "jardín de niños")	{Event, success, gathering}
(comoseusa "se conocieron en una fiesta de cumpleaños")	{child party}
(frasetemat (coloq "para todo haces fiesta" , "para cualquier acto, actividad, acontecimiento se externa al público, amigos, familiares")	{hiponym}
	{Birthday party, baptism party}
	{place}
	{Party room; home; kindergarten}
	{how it is used}
	{They met at a birthday party}
	{thematic phrase}
	{“You make noise for anything”; “any act, activity or happening is exposed to the public, to friends, relatives”}

Table 3. The frame describing “Fiesta Infantil” {children party} showing its elements. Phrases between quotes indicate the description of the corresponding semantic part.

IV. DISAMBIGUATION USING FRAMES

⁴ In this paper, the authors (not the algorithm) have added {English translations} to some inputs and outputs.

Prior to disambiguation, tagging and lemmatizing of words is performed by our method. Our lemmatizer also identifies the root of the word.

4.1 Algorithm of the lemmatizer

It has two steps:

Step 1. It works on an input file which is previously tagged (using the EAGLES [4] standard); this file must have the extension .txt.tag; the tags have the structure shown in Table 4.

Position	Attribute	Value	Code	Position	Attribute	Value	Code
1	Category	Verb	V	5	Person	First	1
2	Type	Main	M			Second	2
		Auxiliar	A			Third	3
		Semiauxiliar	S	6	Number	Singular	S
3	Mode	Indicative	I			Plural	P
		Subjunctive	S	7	Genre	Masculine	M
		Imperative	M			Feminine	F
		Infinitive	N				
		Gerund	G				
		Participle	P				
4	Tense	Present tense	P				
		Imperfect	I				
		Future tense	F				
		Past tense	S				
		Imperfect past	C				
		-	0				

Table 4. Noun, Verb, Adjective tags according to EAGLES standard [Toral et al 2005].

The following example (Table 5) shows the tags given to the sentence “*en ese año se celebró la conferencia de Dartmouth*” {*In that year it was celebrated the Dartmouth conference*}.

Line	palabra {Word}	etiqueta {tag}
1	en {in}	SPS00
2	ese {that}	DD0MS0
3	año {year}	NCMS000
4	se {it}	P0000000

5	celebró {celebrated}	VMIS3S0
6	la {the}	DA0FS0
7	conferencia {conference}	NCFS000
8	de {of}	SPS00
9	dartmouth {Darmouth}	NC00000

Table 5. Words in a sentence, with their tags. The boldface word is the verb which is a candidate to be lemmatized.

Step 2. After tagging, each verb in the sentence is identified, and its corresponding frame is searched. Table 6 shows the frame of “celebrar” {to celebrate}.

(defmarco celebrar (gram "v tr")	{to celebrate} {transitive verb}
(QueEs "Llevar a cabo un acto o ceremonia")	{To cary out an act or ceremony}
(Ejemplo "Se celebraba un mitin en la plaza",	{a meeting was celebrated in the main square}
"Las reuniones se celebraban en su casa",	{Gatherins were held in his home}
"Celebraron una asamblea",	{An assembly was celebrated}
"El sábado se celebraron las elecciones")	{Elections were held on Saturday}
(Raíz " celebr ")	{root}

Table 6 showing a frame with descriptive elements such as *quees* {what is it}, *ejemplo* {example} and *raiz* {root}.

Once the verb is found, it is compared with the root of the frame. Table 6 shows the tag given to *celebró* {celebrated}.

5	celebró	VMIS3S0
---	----------------	---------

Table 7. Comparing the tagged verb with the root in the frame.

If there is coincidence, the verb is recognized and the root identified. It is important to notice here that our disambiguator needs the lemmatizer when it requires to disambiguate using frequent words, for instance “*hablaremos sobre literatura*”, {we will talk about literature} (see tables 20-22); in this case *sobre* {about} is the preposition and the words giving sense to it are *hablaremos* {we will talk} and *literatura* {literature}. Thus, it is required to lemmatize *hablaremos* identifying its root as *hablar* {to speak}. Once done this, our disambiguator will give the sense of the preposition, the result is shown as [theme] {topic} in Table 23.

The Lemmatizer only has those two steps; after it, the document goes to the Disambiguator, which is now explained.

4.2 Disambiguator of propositions.

A hypernym of a word is defined as another word whose meaning contains that of the first word, it is a superclass or superset of it. An hyponym forms part of the morphosyntax of the sentence, that is, the set of elements and rules that allows construction of meaningful sentences. We obtain most of our hypernmys from the **WordNet** semantic data base.

Some examples of hypernmys are given in table 8.

Hypernyms	Derived words
Material {material}	Vidrio, madera, plástico, hule, etc. {glass, wood, plastic, rubber, etc.}
Animal {animal}	Vaca, perro, gato, pato, pez, etc. {cow, dog, cat, fish, etc.}
Persona {person}	Niño, amigo, papá, mamá, etc. {child, friend, father, mother, etc.}

Table 8. Examples of hypernyms.

The lemmatizer receives as input a tagged and lemmatized file. Only words contributing to the meaning of the preposition will be analyzed; these words are the nouns and verbs found before and after of the proposition.

Example: Jarra de Agua {pitcher of water}. *Jarra* and *Agua* are used to disambiguate preposition *de*.

The algorithm often compares the hypernyms of the words preceding and following the preposition.

The flow of the disambiguator is given in Figure 1.

Next, we describe the possible flows or routes that may be followed to disambiguate a proposition, starting with Route 1. A route is the way followed by the algorithm under some conditions.

ROUTE 1.

- The disambiguator identifies the prepositions in the file (which is already tagged and lemmatized).
- Nouns or verbs before and after the preposition are identified. These words contain the information needed to disambiguate the preposition.
- If the word following the preposition is a noun, the field **HiperonimoPost** {following hypernym} of the frame of the noun is compared with the hypernym of the frame of the proposition. If the following word is not a noun, flow continues in **ROUTE 2**.

If there were coincidences when comparing hypernyms with only one frame of the preposition, a meaning is assigned to the preposition and the algorithm ends. If the number of coincides is not one, flows continues to **ROUTE 3**.

ROUTE 2.

- Field “frequent words” of the frame of the proposition is compared with the previous and following words.
- If there are coincidences, a meaning (a sense) is assigned to the proposition and the algorithm ends.

If there are no coincidences in the frequent words and there were no coincidences in the hypernym, the algorithm could not assign a meaning to the proposition, and it ends (with this proposition).

Otherwise, if there is more than one coincidence in the frequent words, more than one meaning is assigned to the preposition, and the algorithm ends.

ROUTE 3.

- If the word previous to the preposition is a noun, the field **HiperonimoAnt** {previous hypernym} of the frame of the noun is compared with the hypernym field of the proposition. If the number of coincidences differs from one, flow goes to **ROUTE 2**, otherwise a sense is assigned to the proposition and the algorithm ends.

If the previous word is not a noun flows goes to **ROUTE 2**.

To show the algorithm in action we will analyze the following sentence:

*“La vida de los **colibríes** está en estrecha relación con la de ciertas plantas y flores; de hecho constituyen una gran importancia para los ecosistemas, pues, son agentes polinizadores de gran cantidad de plantas en los bosques tropicales”.* {Life of hummingbirds is in close relation with that of certain plants and flowers; in fact they constitute a great importance to ecosystems, since they are pollinating agents of a large amount of plants in tropical forests} Table 9 shows the tagged words.

la {the}	DA0FS0	para {for}	SPS00
vida {life}	NCFS000	los {the}	DA0MP0
de {of}	SPS00	ecosistemas {ecosystems}	NCMP000
los {the}	DA0MP0	, {,}	Fc
colibríes {hummingbirds}	NCMP000	pues {since}	CS
esta {is}	DD0FS0	, {,}	Fc
en {in}	SPS00	ser {are}	VSIP3P0
estrecha {close}	AQ0FS0	agentes {agents}	NCMP000
relación {relation}	NCFS000	polinizadores {pollinizers}	NCMP000
con {with}	SPS00	de {of}	SPS00
la {that}	DA0FS0	gran {great}	AQ0CS0
de {of}	SPS00	cantidad {amount}	NCFS000
ciertas {some}	DI0FP0	de {of}	SPS00
plantas {plants}	NCFP000	plantas {plants}	NCFP000
y {and}	CC	en {in}	SPS00
flores {flowers}	NCFP000	los {the}	DA0MP0
, {,}	Fc	bosques {forests}	NCMP000
de {in}	SPS00	tropicales {tropical}	AQ0CP0
hecho {fact}	NCMS000	. {.,}	Fp
constituyen {they constitute}	VMIP3P0		
una	DI0FS0		
gran {great}	AQ0CS0		
importancia {importance}	NCFS000		

Table 9. Tagged words of the sentence, the first column continues in the second column

The disambiguator is formed by the following steps:

Step 1. The algorithm will analyze all the prepositions contained in the fine. Their disambiguation will depend on the resources (frames) available. The prepositions are identified with the tags **SPCMS** y **SPS00** as shown in Table 10.

Palabra	Etiqueta
Para {for}	SPS00
Con {with}	SPS00
En {in}	SPS00
Del {of the}	SPCMS
De {of}	SPS00

Table 10. Prepositions identified in the sentence of Table 9

Step 2. Once prepositions are found, nouns or verbs located before and after each proposition are sought (except when a punctuation mark is found). These are called “significant words” (they provide most of the information used to disambiguate the proposition); their tags start with “N” for a noun and “V” for a verb. These words help to determine the sense of a proposition. Significant words found in this example are shown in Table 11.

Previous noun/verb	Tag	Preposition	Following noun/verb	Tag
vida {life}	NCFS000	de {of}	colibríes {hummingbirds}	NCMP000
relación {relation}	NCMS000	con {with}	plantas {plants}	NCFP000
polinizadores {pollinizers}	NCMP000	de {of}	cantidad {amount}	NCFS000
importancia {importance}	NCFS000	para {for}	ecosistemas {ecosystems}	NCMP000
cantidad {amount}	NCFS000	de {of}	plantas {plants}	NCFP000
plantas {plants}	NCFP000	en {in}	bosques {forests}	NCMP000

Tabla 11. Significant words before and after propositions.

Step 3. Next, the disambiguator looks for the frame of each proposition and the previous significant word. In our example:

“la vida de los colibríes ...” {life of hummingbirds} will look the frame of *colibrí* (all nouns are singular)

- Table 12 shows the frames of preposition “de” with two senses (for lack of space we only show two frames). In our example the preposition corresponds to possession.

(defMarco de(gram "prep")	{of} {preposition}
(QueEs "Significa una relación de posesión o pertenencia, así como de dependencia entre personas o cosas")	{what it is} {Means a relation of possession or belonging, as well as dependence among persons or things}
(Sinonimia "posesión")	{possession, ownership}
(Ejemplo "La casa de mi padre", "Los juguetes de los niños", "Los parques de la ciudad")	{example} {The house of my father; the toys of children; the parks of the city}
(PalabrasFrecuentes "")	{frequent words}
(HiperonimiaPost "persona", "organismo", "organización", "objeto")	{following hypernym} {person; organism; organization; object}
(HiperonimiaAnt "objeto", "organismo", "persona", "estado", "atributo", "órgano")	{previous hyperym} {object; organism; person; state; attribute; organ}
)	
(defMarco de (gram "prep")	{of} {preposition}
(QueEs "Indica el todo o el conjunto Del que se torna o separa una parte")	{what it is} {It indicates the whole or set of which a part is taken or separated}
(Sinonimia "grupo")	{synonymy} {group}
(Ejemplo "Dos de ellos ganaron", "Pagué parte de la deuda")	{example} {"Two of them won;" " I paid part of the debt"}
(PalabrasFrecuentes "grupo")	{frequent words} {group}
(HiperonimiaPost "persona", "organismo", "organización")	{following hypernym} {person, organism, organization}
(HiperonimiaAnt "cantidad", "grupo", "magnitud")	{previous hypernym} {amount, group, organization, magnitude}
)	

Table 12. Two frames of proposition “de”.

Table 13 shows a frame of noun “colibrí”.

(defMarco colibrí(gram "s m sing")	{hummingbird} {singular masculine}
(QueEs "pájaro de la familia de los troquélicos, exclusiva de américa, de aproximadamente 7 cm de longitud, de pico largo y delgado, con un plumaje muy brillante y vistoso en el que predomina el color verde.")	{what it is} {Bird of the troquelidos family, exclusive of America, of approximately 7 cm length, long and thin beak, with very bright and colorful plumage, where green color dominates}
(Hiperonimia "ave", "animal", "vertebrado", "organismo")	{hypernym} {bird, animal, vertebrate, organism}
)	

Tabla 13. Frame of *colibri*, to save space only some semantic fields are shown

Step 4.

The disambiguator will compare the hypernyms (general description of a concept) of all frames of the preposition under analysis (Table 12) against the hypernyms of the frames of the following significant word (Table 13). In our example the following hypernyms are compared (Table 14), all of this in order to identify the context of the preposition.

Hypernyms of “de”	Hypernyms of “Colibrí”
Person	bird
organism	animal
organization	vertebrate
object	organism

Table 14. Comparing hyperms (first sense of “de”).

Now we compare the hypernyms of the second sense of the preposition. See Table 15.

Hypernyms of “de”	Hypernyms of “Colibrí”
person	bird
organism	animal
organization	vertebrate
	organism

Table 15. Comparing the hypernyms of the second sense of “de”.

If coincidences are found between the hypernym of just only one frame of the preposition with some hypernym of the following significant word, the meaning of the proposition is that of the field “synonymy” of the coinciding frame (of the preposition) and the algorithm ends.

In our example exist coincidence between the hypernms of the following noun and two frames of proposition “de” (Tables 14 and 15).

In this case where more than one frames coincide, the former significant word is analyzed. It is *vida* {life}. Table 16 shows the frame of *vida*.

(defMarco vida(gram "s m sing")	{life}
(QueEs "estado de actividad de los seres	{what it is}
orgánicos por el que se desarrollan,	
evolucionan y se reproducen.")	{State of activity of organic beings by which they
	grow, evolve and reproduce }
(Hiperonimia "existencia","estado",	{hypernym}
"atributo","abstracción")	{Existence, state, attribute, abstraction}
)	
(defMarco vida(gram "s m sing")	{life}
(QueEs "modo de pasar este tiempo las	{what it is}
personas, según su actividad, su conducta,	{Way by which people spend time, according
la satisfacción de sus necesidades y de sus	to their activity, their behavior, the satisfaction
deseos")	of their needs and of their wishes }
(Hiperonimia "experiencia","conocimiento",	{hypernym}
"rasgo psicológico","abstracción")	{Experience, knowledge, psychological
	trait, abstraction }
)	

Table 16. Frames of concept *vida*.

The hypernyms of the former significant word (field **HiperionimiaAnt**) are compared against the hypernyms of each frame of the proposition (Table 12). Table 17 shows the comparisons against the first sense of the proposition.

Hypernyms of “de”	Hypernyms of “vida”	Hypernyms of “vida”
object	existence	experience
organism	state	knowledge
person	attribute	Psychological trait
state	abstraction	abstraction
attribute		
organ		

Table 17. Comparison of the three intervening hypernyms. **attribute** of hypernym “de” (first column) coincides with **attribute** of hypernym of “vida” (as existence), middle column.

Now, hypernyms of the second sense of the preposition are compared. There are no coincidences. See Table 18.

Hypernyms of “de”	Hypernyms of “vida”	Hypernyms of “vida”
quantity	existence	experience
group	state	knowledge
magnitude	attribute	Psychological trait
	abstraction	abstraction

Table 18. Comparing the hypernyms of “de” (second sense) against the two senses of concept *vida*.

As we can see, only there is coincidence of the hypernyms of the first sense of the preposition against the hypernyms of the previous noun. Thus, the sense of the preposition is the field “synonymy” of the coinciding frame; thus, the meaning is *posesión* {*possession, ownership*}. See table 19.

la	DA0FS0	
vida	NCFS000	
de	SPS00	[posesión]
los	DA0MP0	
colibríes	NCMP000	

Table 19. Output showing disambiguation of “de” as *posesión* {ownership, possession}.

Step 5. This step is carried out if the significant word following the proposition is not a noun, or if a sense has not been determined in the former steps.

a) The field **PalabrasFrecuentes** {frequent words} of the frames of the proposition is compared with the previous and following significant words around the proposition. An example with the sentence “hablaremos sobre literatura...” {let us talk about literature...} shows some frames of the proposition “sobre” {above; envelope; about; upon} in Table 20.

(defMarco sobre(gram "prep")	{over, above}
(QueEs "Indica que algo o alguien está situado en un nivel o en una jerarquía superior a la de otra cosa o persona")	{what it is}
	{It indicates that something or somebody is situated in a level or in a hierarchy superior to the other thing or person}
(Sinonimia "jerarquía")	{hierarchy}
(Ejemplo "El presidente está sobre los ministros")	{The President is above the ministers}
(PalabrasFrecuentes "persona")	{person}
(HiperonimiaPost "persona", "organización")	{person, organization}
(HiperonimiaAnt "persona", "organización")	{person, organization}
)	
(defMarco sobre(gram "prep")	{about}
(QueEs "Introduce el tema o el asunto de que trata algo")	{what it is}
	{It introduces the theme or subject of which something talks}
(Sinonimia "tema")	{synonymy} {topic, theme}
(Ejemplo "Habló sobre sus experiencias didácticas")	{example}
	{He talked about his didactic experiences}
(PalabrasFrecuentes "interactuar")	{frequent words} {interact}
(HiperonimiaPost "ciencia", "disciplina", "organismo")	{following hypernym}
	{science, discipline, organism}
(HiperonimiaAnt "interactuar")	{previous hypernym} {interact}
)	

Table 20. Frames of the proposition *sobre*; only two different senses are shown.

Table 21 shows the frequent words related to the first sense of preposition “sobre”.

Frequent words	Previous and following significant words
persona {person}	hablar {to talk}
	literatura {literature}

Table 21. The first column shows the frequent words of *sobre* taken from its first sense (hierarchy).

For the second sense (second frame) of *sobre* (Table 20) we have Table 22.

Frequent words	Previous and following significant words
interactuar {to interact}	hablar {to talk}
	literatura {literature}

Table 22. The first column shows the frequent words of *sobre* taken from its second sense (topic).

With the first frequent word of the frame, “person” (Table 21), the algorithm consults in a dictionary the words related to it. This dictionary contains groups of words related to the encompassing word; it is part of the linguistic resources used by the algorithm. Its contents are created from the WordNet data base. Thus, the algorithm finds that “hablar” {to talk} and “literatura” {literature}, do not belong to the set of words related to *person*. See Figure 2.

7	@persona	{person}
8	papá	{father}
9	mamá	{mother}
10	presidente	{president}
11	ministro	{minister}
12	gobernador	{governor}
13	hijo	{son}
14	hija	{daughter}

Figure 2. Words in the dictionary expanding the concept *persona* {person}.

If we consider the frequent words of the second frame of preposition *sobre* (table 20), the algorithm determines that word *hablar* {to talk} is in the set of words related to word *interactuar* {to interact}, as shown in Figure 3.

1	@interactuar	{to interact}
2	platicar	{to chat}
3	hablar	{to talk}
4	escribir	{to write}
5	preguntar	{to ask}
6		

Figure 3. List of words in the dictionary expanding the concept *interactuar* {to interact}.

Thus, it is determined that the sense of *sobre* is the second sense (table 20). Its output is shown in Table 23.

hablar	VMN0000	{to talk}
sobre	SPS00	[tema] {topic}
literatura	NCFS000	{literature}

Table 23. Preposition disambiguated with the sense *tema* {theme, topic}. The algorithm adds [tema] to the output.

If no coincidence in frequent words is found, the algorithm is unable to determine the meaning of the proposition; it could be improved adding more frequent words to the dictionary.

If significant words coincide with just one frame, the meaning of the proposition is that of the field “synonymy” of the frame; if there are more than one coincidence, the algorithm determines more than one sense for the proposition, for instance *procedencia* {place of origin} and *parte* {part of}. Table 24 shows this.

científicos	NCMP000	{scientists}
de	SPS00	[procedencia, parte]
méxico	NCMS000	{place of origin; part of}

Table 24. Where a proposition has been identified with two senses: *procedencia* {place of origin} and *parte* {part of}. The algorithm adds [procedencia, parte] to the output. It narrowed down the meaning of *de* but not completely.

V. RESULTS

Examples of preposition “DE”.

Input is a file with extension .txt.tag.

The input to the disambiguator is shown in Table 25.

	divisas	NCFP000	{currency}
	de	SPS00	
	europa	NCFS000	{Europe}
*	.	Fp	
	programa	NCMS000	{program}
	de	SPS00	
	computadora	NCFS000	{computer}
*	.	Fp	

Table 25. The list of tagged words, candidates for disambiguation

Figure 4 shows the output of the disambiguator.

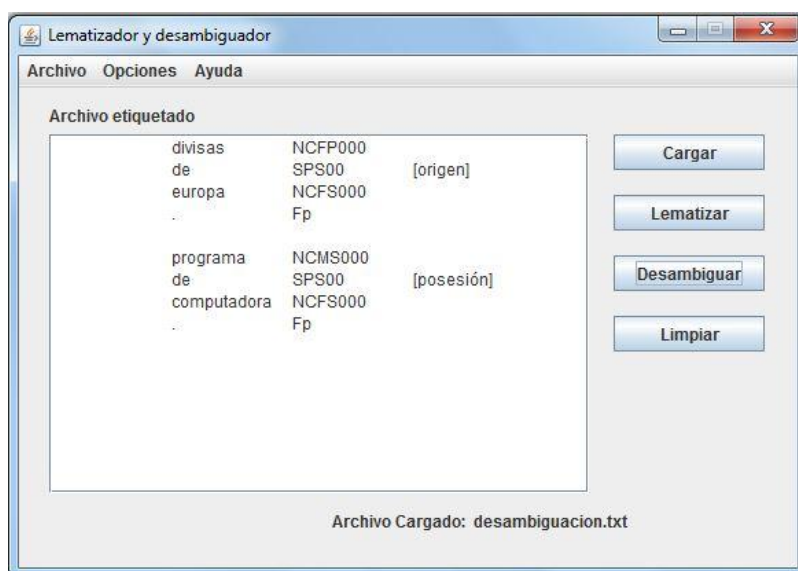


Figure 4. Output screen showing the proposition *de* with two meanings, the first one is *origen* {procedence, origin} in *divisas de Europa* {currencies from Europe}, the second sense is *posesión* {ownership}, in *programa de computadora* {computer program}.

Examples for proposition “PARA”.

Table 26 shows the list of tagged words.

Ejemplos	{examples}	NCMP000	mis	{my}	DP1CPS
para	{for}	SPS00	trabajos	{works}	NCMP000
desambiguar	{to disambiguate}	VMN0000	con	{with}	SPS00
*.	{.}	Fp	la	{the}	DA0FS0
este	{this}	DD0MS0	escuela	{school}	NCFS000
documento	{document}	NCMS000	*.	{.}	Fp
ser	{to be; is}	VSIP3S0	necesitar	{to need}	VMIP1S0
para	{for}	SPS00	llegar	{to arrive}	VMN0000
la	{the}	DA0FS0	para	{for}	SPS00
secretaría	{Secretary, Ministry}	NCFS000	el	{the}	DA0MS0
*.	{.}	Fp	anoecer	{nightfall}	NCMS000
salgo	{I leave}	VMIP1S0	*.	{.}	Fp

Figure 6. Disambiguated sense for proposition *para* {for} is *uso* {use} in *ser (es) malo para el viaje* {it is bad for the journey} and *plazo* {deadline, period} in *necesitar (necesito) llegar para el anochecer* {I need to arrive at nightfall}.

Figura 7 shows disambiguation of *para* in phrase *ser (es) malo el día para el viaje* {the day is bad for the journey}.

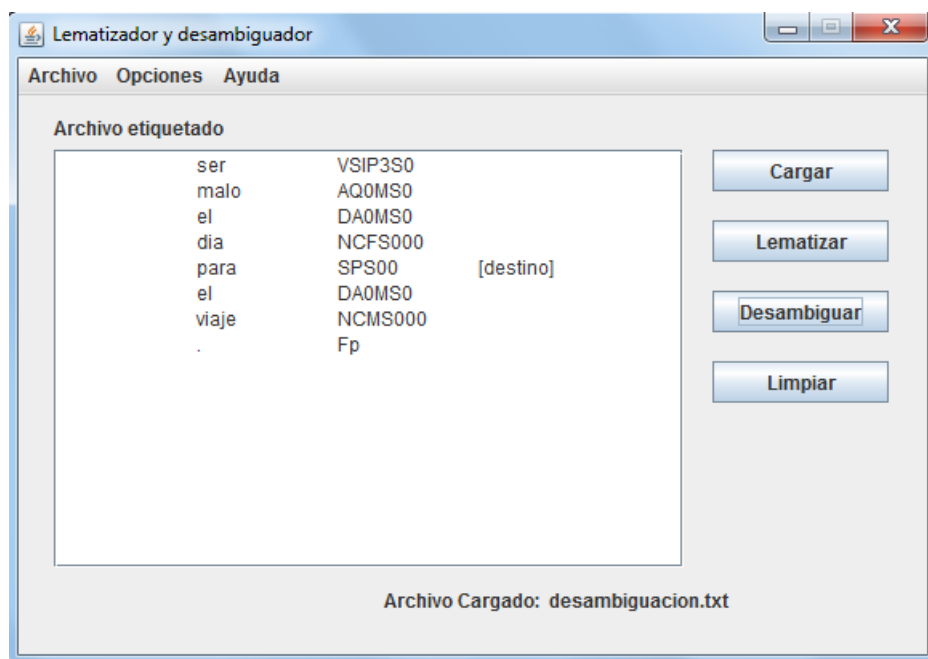


Figure 7. Disambiguated output where the sense of proposition *para* is given as *destino* {destination}.

Examples for proposition “EN”.

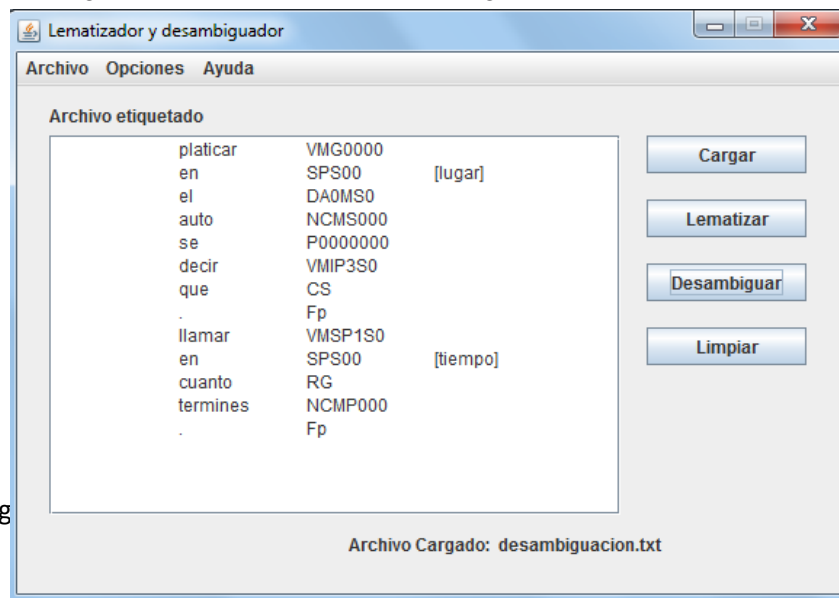
Table 27 gives a brief example, with short sentences.

platicar	{to talk}	VMG0000	la	{the}	DA0FS0
en	{in}	SPS00	fiesta	{party}	NCFS000
el	{the}	DA0MS0	ser	{to be}	VSIP3S0
auto	{car}	NCMS000	mejor	{better}	AQ0CS0

se	{is}	P0000000	en	{in}	SPS00
decir	{to say}	VMIP3S0	compañía	{company}	NCFS000
que	{that}	CS	de	{of}	SPS00
.		Fp	los	{the}	DA0MP0
llamar	{to call}	VMSP1S0	amigos	{friends}	NCMP000
en	{in}	SPS00	.		Fp
cuanto	{soon --as soon as}	RG	el	{the}	DA0MS0
termines	{you finish}	NCMP000	hierro	{iron}	NCMS000
.		Fp	se	{is}	P0300000
la	{the}	DA0FS0	funde	{to melt}	VMIP3S0
casa	{house}	NCFS000	en	{in}	SPS00
de	{of}	SPS00	caliente	{hot}	AQ0CS0
juárez		NCFS000	.		Fp
ir	{to go}	VSIS3S0			
construida	{built}	VMP00SF			
en	{in}	SPS00			
1806		Z			
.		Fp			

Table 27. Each word shows its tag.

Output of the disambiguator for Table 27 is shown in Figure 8:



Disambiguating

Figure 8. Output of the algorithm, where proposition *en* has the meanings *lugar* {place} in *platicar en el auto* {to talk in the car} and *tiempo* {time} in *llamar en cuanto termines* {call as soon as you finish}.

Figure 9 presents output for the sense of preposition *de* as *posesión* {ownership} in *la casa de Juárez* {the house of Juárez} and for preposition *en* as *tiempo* {time} in *construida en 1806* {built in 1806}.

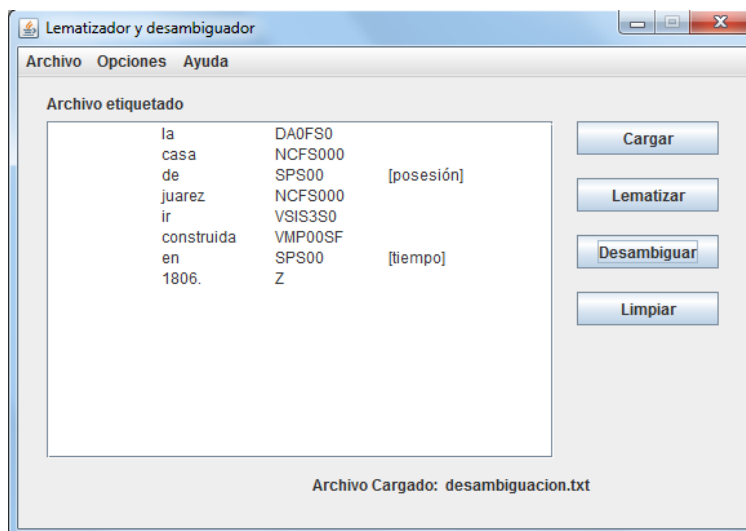


Figure 9. The algorithm assigns meanings to propositions *de* and *en*.

Remark:

In the following sentences:

- La fiesta es mejor *en* compañía de los amigos {the party is better in company of friends}
- El hierro se funde *en* caliente {Iron melts when hot}

the disambiguator assigned no meaning to preposition *en*. In the second sentence, the word *caliente* {hot} is not a noun, thus the algorithm did not take it into account.

Examples of disambiguation of preposition “a”.

Consider the following sentences:

1. Recibo a los invitados {I receive the guests}
2. Me iré a Oaxaca {I will go to Oaxaca}
3. El trabajo está destinado a fracasar {work is destined to fail}

Tagging of those sentences is shown in table 28.

recibir	VMIS3S0	{to receive}
a	SPS00	{to}
los	DA0MP0	{the}
invitados	NCMP000	{guests}
.	Fp	
me	P010S000	{I}
ir	VMIP1S0	{to go}
a	SPS00	{to}
oaxaca	NCFS000	
.	Fp	
el	DA0MS0	{the}
trabajo	NCMS000	{work}
estar	VMIP3S0	{to be}
destinado	AQ0MSP	{destined}
a	SPS00	{to}
fracasar	VMN0000	{to fail}
.	Fp	

Tabla 28. Tagged sentences.

Output of the disambiguator is shown in Figure 9.

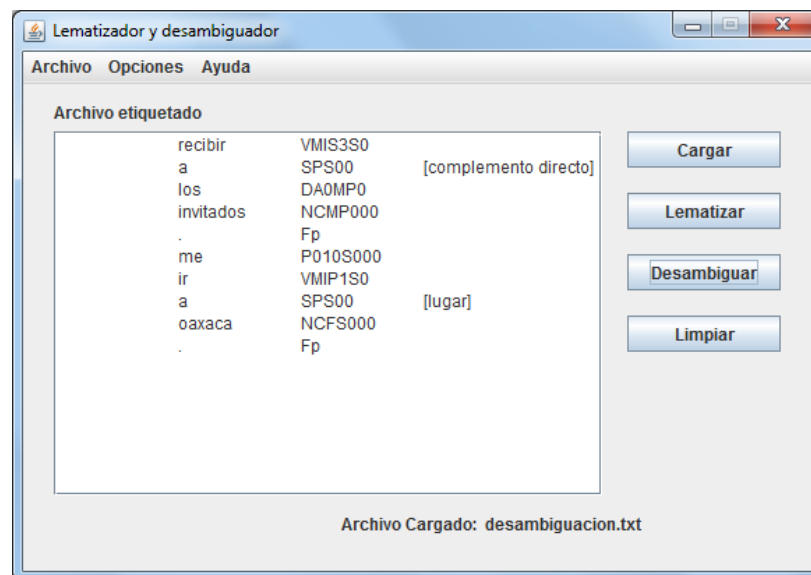


Figure 9 presenting two senses for proposition *a* {to}, as *complemento directo* {direct complement} and as an adverb of *lugar* {place}.

Figure 10 contains additional output.

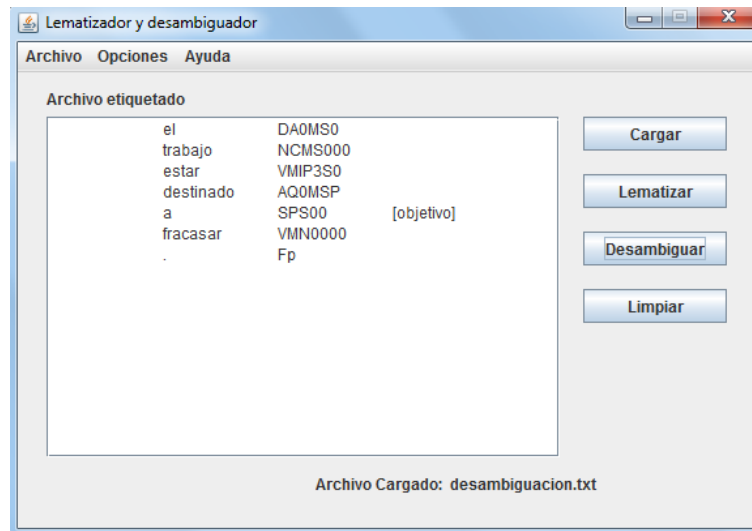


Figure 10 gives to proposition *a* {to} the meaning *objetivo* {objective} in *el trabajo está destinado a fracasar* {work is destined to fail}.

To quantify the results given by the disambiguator, we compute the percentage of correct answers for each proposition with the formula:

$$P = (T * 100) / PA$$

where:

P= percentage of correct answers (%)

T= Total of correct answers

PA=Prepositions to be tested.

Preposition	Analyzed	Correct	Wrong	Porcentaje (%)
A	7	5	2	71.42
de	28	25	3	89.28
con	7	7	0	100
según	4	4	0	100
para	223	18	0	91.92
sin	92	91	1	98.91

Table 29. Results for our disambiguator.

V. CONCLUSIONS

The disambiguator starts with a list of Spanish prepositions to be disambiguated, with their different meanings. The lemmatizer shown is based in the frames of the prepositions, as used in Spanish, in order to tag the nouns and verbs and the hypernoms of prepositions and nouns of the sentence. This disambiguation technique, using frames, has not been used before. Our algorithm is unique in this regard; it is also a heuristic algorithm. To better assess its efficiency a larger set of frames should be defined (constructed).

The disambiguator obtains good results (Table 29). The use of hypernoms is crucial.

ACKNOWLEDGMENTS

To ESCOM-IPN, where one of the authors (D. M) presented her thesis #2010-0069, which gives a more detailed description of the disambiguator. Work herein reported was partially sponsored by Conacyt Grant #128163, by IPN (A.G. as Resident Scientist) and by SNI.

REFERENCES

- Allan, M. Collins, A; M.R. Quillian (1969). [Retrieval time from semantic memory](#). *Journal of verbal learning and verbal behavior* **8** (2): 240–247. doi:10.1016/S0022-5371(69)80069-1
- Allen, J. (1987) [Natural language understanding](#). California: The Benjamin/Cummings Publishing Company.
- Colorado F. (2008) [Mapeando palabras a conceptos](#). Tesis de Maestría. CIC-IPN.
- Cuevas, A. (2006) [Ontology union using semantic properties](#). Ph. D. thesis, CIC-IPN. In Spanish.
- Galicia, H. (2000) [Análisis sintáctico conducido por un diccionario de patrones de manejo sintáctico para lenguaje español](#) Ph. D. thesis, CIC-IPN. In Spanish www.gelbukh.com/thesis/Sofia%20Galicia%20Haro%20-%20PhD.pdf
- F. Vasilescu, P. Langlais, G. Lapalme (2004) [Evaluating variants of the Lesk approach for disambiguating words](#), LREC.
- Hausser, R. (2001) [Foundations of computational linguistics: Human_computer communication in natural language](#), Berlin: Springer.
- Litkowski K. and Hargraves O. (2007) [Task 06:Word-Sense Disambiguation of Prepositions](#). in *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*. Prague, Czech Republic, Association for Computational Linguistics, pp. 24-9.
- McCarthy, D. (2003). [Word Sense Disambiguation: The Case for Combinations of Knowledge Sources](#), by Mark Stevenson. CLSI, 2003.
- Minsky, M. (1975) [A Framework for Representing Knowledge](#). MIT-AI Laborarorio Memo 306, Junio 1974 Reprinted in *The Psychology of Computer Vision*, P. Winston, ed. McGrawHill, 1975. Link <http://web.media.mit.edu/~minsky/papers/Frames/frames.html>
- Molina, A. (2004) [Desambiguación en procesamiento del lenguaje natural mediante técnicas de aprendizaje automático](#). PhD tesis, Universidad Politécnica de Valencia Valencia, Spain.

Mooney, Raymond J. (2003) [Fundamentals, Part I ch. II, III, IV, V](#). *Oxford Handbook of Computational Linguistics*, Oxford University Press. (Ruslan Mitkov Ed.).

Moreno, L., Palomar, M., Molina, A., y Fernández, A. (1999) [Introducción al Procesamiento del Lenguaje Natural](#). (Ed. Servicio de Publicaciones Universidad de Alicante). Universidad de Alicante.

Palomino K., Rosero R., Zapata C. (2007) [Method for Coordinative and Prepositional Syntactic Disambiguation](#). Universidad Nacional de Colombia. Redalyc. Sistema de Información Científica. Red de Revistas Científicas de América Latina, el Caribe, España y Portugal. Mayo 2007. ISSN (Printed version): 0012-7353

R. Mihalcea y T. Pedersen, [Advances in Word Sense Disambiguation](#), Notes of tutorial at AAAI, 2005.

Nübel R. (1996) [Knowledge sources for the Disambiguation of Prepositions in Machine Translation](#). *Proceedings of the Workshop on Future Issues for Multilingual Text Processing*, Pacific Rim International Conference on Artificial Intelligence, Springer, Berlin,

Suárez, Cueto A. (2004) [Resolución de la ambigüedad semántica de las palabras mediante modelos de probabilidad de máxima entropía](#) PhD Tesis, Universidad de Alicante. Alicante, Spain.

Toral A., Ferrández S., Montoyo A. (2005) [EAGLES compliant tagset for the morphosyntactic tagging of Esperanto](#). In *Proceedings of the 5th International Conference on Recent Advances in Natural Language Processing (RANLP)*. Borovets, Bulgaria.

Zapata, C., Arango, F. [Los modelos verbales en lenguaje natural y su utilización en la elaboración de esquemas conceptuales para el desarrollo de Software: Una revisión crítica](#). Revista *Universidad EAFIT*. Vol. **41**. Pp 77-95.

Referencias de Internet

Web site 1 <http://www.lsi.upc.edu/~nlp/tools/parole-sp.html> . Consultado 11 de abril de 2012.

Web site 2 Natural Language Toolkit. <http://www.nltk.org/> [Consultado 11 de abril de 2012].

Web Site 3 Lista de Preposiciones http://www.escueladigital.com.uy/espaniol/8_preposic.htm [Citado 11 de abril de 2012].

Web site 4 Natural Language Toolkit. <http://nltk.sourceforge.net/> [Citado 29 de mayo de 2012].

Web site 5 The Python Programming Language. <http://www.python.org/> [Citado 29 de mayo de 2012].