Hindawi
Mathematical Problems in Engineering
Volume 2018, Article ID 7086965, 14 pages
https://doi.org/10.1155/2018/7086965

Research Article

ANACONJ Analyzer of the Conjunction AND in Spanish Using Syntactic Patterns and Semantic Frames

Alma Delia Cuevas Rasgado

Universidad Autónoma del Estado de México Centro Universitario Texcoco, Av. Jardín Zumpango S/N, El Tejocote, 56259 Texcoco, MEX, Mexico

Correspondence should be addressed to Alma Delia Cuevas Rasgado; almadeliacuevas@gmail.com

Received 25 August 2017; Revised 13 April 2018; Accepted 8 May 2018; Published 20 June 2018

Academic Editor: Vitaly Kober

Copyright © 2018 Alma Delia Cuevas Rasgado. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conjunctions have different interpretations: they eliminate redundancies: "María se bañó y se peinó" (*Maria bathed and she combed her hair*), unite different ideas: "Hoy llovió y no fui a corer" (*Today it rained and I did not go to run*) and use of lists: "Eran Alma, Edith y Omar" (*They were Alma, Edith and Omar*). Conjunctions take different semantic contexts. We understand each other because of the common sense despite expressing ourselves incorrectly from the standpoint of semantics, but for a computer it is difficult. In order to "understand" the sentences, the machine must solve semantics problems; this article exposes one of these problems. ANACONJ is an algorithm of pattern recognition of texts, which uses rules and syntactic patterns that analyze each word of a sentence in a phrase, identifying those sentences with conjunctions to build a semantic tree of the sentence where the conjunction connects words (nouns, verbs, etc.) according to their meaning. ANACONJ could be used as a teaching Spanish software tool and as an app for a service robot too.

1. Introduction

Making computers "interpret" the right semantic of different expressions as complete sentences in which it includes instructions used in a regular language spoken and written is not a trivial task. Natural language has an inherent characteristic: the ambiguity, which the human being solves with the context, based on the real world experience and the common sense. From those aspects that computer just uses the context to identify the meaning of the written words. That is even more complex for developing the app based on the common sense is not part of the computer programming. There are several problems that the machine must solve such as problems of disambiguation of prepositions [1] and of nouns [2], resolution of anaphoric references [3], solution of indirect anaphora [4], and finally the semantic analysis. The way we solve the problem of identifying the meanings of a conjunction is by using syntactic rules and patterns, in addition to semantic frames (we are solving ambiguity with two important tools in information processing).

To get the right text's semantics, it is necessary to use the grammar rules and adapt them to the machine programming language [5]. The Natural Language Processing (NLP) is the responsible area of processing information in common language [6], and in this area there are several techniques which help the computer to process information such as the following:

(i) Morphological analysis (tagging)

(ii) Syntactic analysis (parsing)

(iii) Word-sense disambiguation

(iv) Semantic analysis

Semantic frameworks are fragments of data structures that represent a scenario, for example, the description of the concept *restaurant* unfolds on the stage where commands are performed, menus are observed, accounts are paid, customers are registered, there is furniture such as chairs, tables, and lamps, and there are people who play a role as waiters, head of servers, cooks, diners, and so forth. All these elements (which we call concepts) are located and described in the semantic framework. Particularly we connect the semantic frameworks in an ontology.

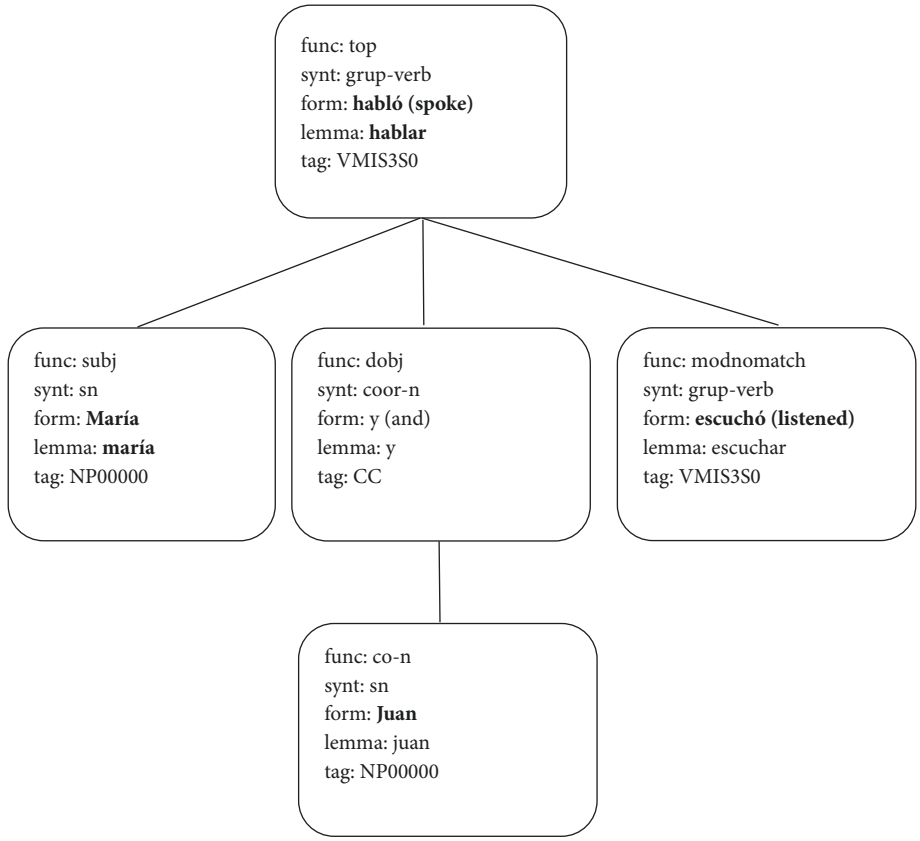


Figure 1: The tree of dependences generated by the FreeLing analyzer from the phrase "Mary spoke and John listened to her."

The ontology as an etymological root that comes from Latin ONTO (being) and LOGOS (treaty) which is the treatise of being and in turn has two meanings: (a) the philosophical and (b) the computational, the first intervenes with the ontology of being, as a deep meaning of knowledge, experience, and praxis of a person, and in the second it refers to the knowledge base that stores everything about a set of objects, ideas, and events, which are related to each other according to their meaning. For example, the concept of a *cat* whose meaning refers to a quadruped animal with hairs and ears and that maul is closer semantically to hairs, play, and milk than a mechanical object, change wheels, vehicle, etc.

The analysis of conjunctions presented in this paper applies to any descriptive text (written in prose describing events, textbooks, recipes, tools, etc.) to represent it in a structure according to the computer. The use of NLP, information retrieval, and knowledge acquisition tools has made some progress in the analysis, mainly in English documents, but for Spanish there are a few. This article represents a contribution to the NLP for Spanish that could be implemented in a service robot such as SABINA (http://ccc.inaoep.mx/busqueda.php?cof=FORID:11&cx=004801418483011671-169:-8q4-3zbk5y&q=sabina) that receives indications in natural language and is transported in a housing space to carry out the action that is indicated to the robot.

1.1. Problem Statement. The conjunction can be analyzed at language level; however the tasks oriented to the computer science allow us to demonstrate these theoretical analyses through algorithms. Today there are many computational applications that solve problems of the natural language where there are the problems of pattern recognition in texts. These applications do not analyze the semantic context; in other words ANACONJ is able to interpret whole sentences, not isolated words, including the context as well based on semantic and the applied linguistics.

One of the applications that parses sentences in natural language is FreeLing [7] and since its analysis is purely grammatical, when analyzing sentences that have conjunctions there are some details that deserve to be discussed, and the analyzer does not recognize the sense of the conjunction, within the sentence. For example, the text "María habló y Juan escuchó" (*Mary spoke and John listened to her*) indicates that the conjunction *and* can be used in different senses, for example, (1) to join two ideas "*Mary spoke and John listened to her*" and (2) to indicate the last element of a list "*we need to buy milk, bread, ham and eggs*." In the first case "*Mary spoke and John listened to her*," there are two ideas, each one conformed by a verb and a noun phrase, first "*Mary spoke*," and second "*John listened to her*." Two separate ideas could be considered; that is, they do not have a semantic relation between them, so that their nodes (represented by rectangles in Figure 1) should not be linked in the tree of dependencies.

A dependency tree is a set of nodes that are related through their grammatical relationship. It can be observed in Figure 1 that the verb *spoke* is united with the verb

VESTIMENTA *{clothing}*	PREFERIBLE INFORMAL*{informal is better}*
REGALO *{gift}* ----------	DEBE COMPLACER AL ANFITRIÓN *{it should like the host}*
	DEBE SER COMPRADO Y ENVUELTO *{it must be bought and wrapped}*
JUEGOS *{games}* --------	ESCONDIDILLAS, PONLE LA COLA AL BURRO *{hide-and-seek, Pin the tail on thedonkey}*
DECORACIÓN *{decor}*--	GLOBOS, SERPENTINA, PAPEL CREPÉ *{toy balloons, serpentine streamer, trimmings}*
COMIDA DE FIESTA *{party food}*	PASTEL, HELADO, REFRESCO, HOT DOGS *{cake, ice cream, soda, hotdogs}*
PASTEL *{cake}* ----------	VELAS, SOPLAR, DESEO, CANTAR CANCIÓN DE CUMPEAÑOS *{candles, to blow, wishes, to sing Happy birthday to you}*
HELADO *{ice cream}*--	ESTÁNDAR, TRES-SABORES *{one flavor, three flavors}*

Figure 2: Frame depicting a children's birthday party.

listened and also unites the conjunction *and* with John, which semantically does not apply in reality.

The way in which the tree nodes are related semantically is identifying the meaning of the conjunction in the sentence using semantic frames (to be explained in Section 2). ANACONJ represents a module of treatment of conjunctions that identifies the function that fulfills the conjunction within the sentence. According to the analysis of semantics, the result returned by ANACONJ represents the grammatical and semantic relationship between the nodes of the tree, which is often close to the way of human's relation.

1.2. Challenges. The PNL has greatly progressed for texts in English (many works in English or German mention that his work is expandable, efficient, and independent of language, or "applicable with minor and trivial changes". We have seen that these changes are heavy and in general have not been tested) but not for Spanish. Many tools and resources are still missing; for example, the EAGLES standard [8] is available, but a semantic tagger is needed to help identify the semantic closeness between the elements of a sentence. However, ANACONJ uses the EAGLES standard and is based on semantic frameworks that together form a semantic network or ontology to solve the problem of the identification of meanings.

1.2.1. Contributions.

(i) Automatic identification (unsupervised) of meanings in conjunctions in natural language texts.

(ii) Use of frames [9] through OM (Ontology Merging) notation [10] that provides more detail to descriptions of frame. Its notation was created by the author of ANACONJ.

(iii) The fact that ANACONJ contributes to the simple representation of a text; for example, it is useful as a didactic tool in the process of learning Spanish in the basic and high levels. Also it is very useful in the learning of the language by foreign people and to specify the meaning of an order received by a service robot, for example, SABINA, since if the order is ambiguous it increases the percentage of error of the robot.

2. Semantic Frames to Represent the Knowledge

This is a partial theory of thought that combines a number of classical and modern concepts of psychology, linguistics, and artificial intelligence. The essence of the theory is the following. When a person discovers a new situation (or makes a substantial change in the way of seeing a situation) the person selects from his memory a structure called Frame; this is an agreed framework that adapts to fit with reality, changing details as required. This implies that the frames keep the knowledge that a person has about a given situation.

A frame is a data structure to represent a stereotyped situation, for example, meeting in a certain type of room, or going to a children's birthday party. Each frame has different types of information about how to use the frame, what you can expect to happen in that context, and what does not occur in it.

The meaning for a child's birthday party is very poorly approximated to the definition of any dictionary, such as "a party held to celebrate a birthday" where a party can be defined, in turn, as "people gathered together for a celebration." It has not all the detail of the activities carried out in it. Children know that the "definition" should include more specifications. Figure 2 presents the frame or scenario with the activities and characteristics of the definition of children's birthday party.

These elements for a typical American birthday party should be established over a period of time. Larger events take place on one or more days. A party takes place in a day, of course, and occupies a substantial part of it, and then it is located in its respective day frame.

In the following frame (Figure 3) the noun "institution" is described in the OM notation (to be explained in Section 3.2). The meaning of each of the elements that make up the frames will be mentioned later, although we can use role = node to mean that the concept *institution* is a main node of the frame. A set of frames forms an ontology.

Figure 4 shows the example of a frame describing a verb (hold), where agent indicates who realizes the action of "hold," passive indicates who is affected by the action, *cosa-viviente* (*living thing*), and *objeto-artificial* (*artificial object*) are categories called active agent and passive agent. For example, person has the category: *living thing*, and bottle has

```
<concept> Institución
        <Language>Spanish
                <word> institución, establecimiento</word>
                </gloss> Organización fundada y reunida por un propósito en específico</gloss>
        </Language>
        <subset> organización </subset>
        </relation>role = nodo </relation>
</concept>
```

Figure 3: Frame that describes the noun "institution."

```
<concept> sujetar
        <Language>spanish
                <word>atrapar, sujetar</word>
                <gloss>El acto de sujetar un objeto con las manos</gloss>
        </Language>
        <subset> tomar</subset>
        <relation>role = relation </relation>
        <relation>agente = cosa-viviente, objeto-artificial </relation>
        <relation>pasivo = cosa-viviente, objeto-artificial </relation>
</concept>
```

Figure 4: Frame that describes the verb "Hold."

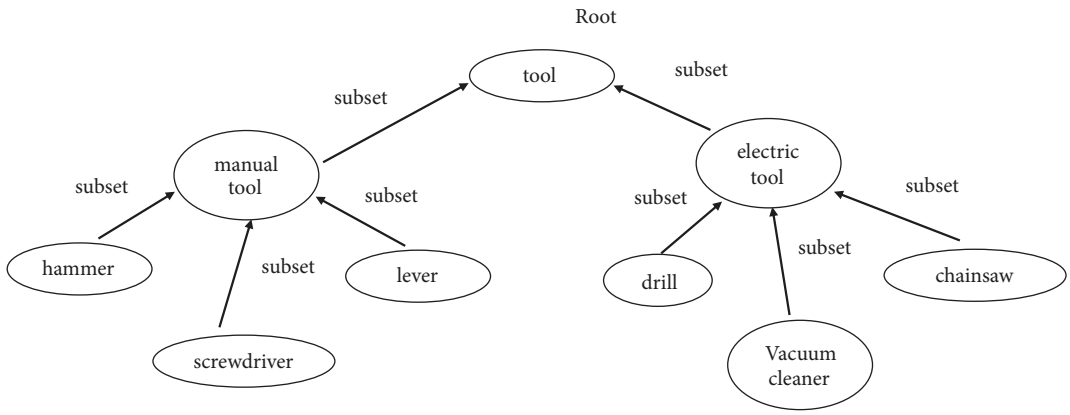


Figure 5: Example of ontology called "tool."

the category *artificial object* (that is to say, is not natural thing). In conclusion, hold is a verb that can be used between two words: person holds a bottle; dog holds the ball.

3. Theoretical Framework

ANACONJ uses an ontology, and for that reason a definition is presented below.

3.1. Ontology. Something that is important to mention is the relationship between ontology and semantic frameworks. Frames serve the ontology to identify the relationship that exists between the concepts that compose it. This is because each of the concepts stored in the ontology is described by a semantic framework, which indicates different characteristics of the object.

ANACONJ makes use of ontologies. An ontology in general is defined as "a representation or formal specification of a shared conceptualization" [11].

An ontology is like a great tree with its branches (nodes), including the branches and the baby branches coming out from them, excepting the root of the tree. For example, suppose you have a root called *tool*, and then you will have as baby node a *manual tool* and an *electric tool*. *Manual tool* will in turn be the father node of *hammer*, *screwdriver,* and *lever* nodes while *electric tool* will be the parent of all electric tools: *drill*, *vacuum cleaner*, *chainsaw*, and nodes. See Figure 5.

The composition of ontologies, similar to a tree, allows us to traverse them in a simple way to identify the relationship between the concepts that compose it. It is also necessary to take into account the fact that all the concepts in the ontology are related; that is to say, all the concepts in the ontology will be linked to the same root node, the reason why one has to be careful when implementing methods to cross that ontology, since if you go through the ontology passing through all the parents of a concept, you will come to that root. Figure 5 shows a graphic example of how an ontology is structured.

The accuracy of the results obtained in ANACONJ depends largely on the correct design of the ontology, i.e.,

according to reality, taking care of the correct representation of homonyms (*hammer*, *screwdriver*, and *lever*) to avoid ambiguities, identifying the concept sense in the ontology. Those who develop the ontology sometimes do not know all the meanings of a concept or do not design the ontology according to norms and standards. The standard that has been followed for the construction of the ontology of ANACONJ is the one that indicates the dictionary of the RAS (Royal Spanish Academy).

3.2. The OM Language. The OM Notation [10] represents ontologies through XML (eXtensible Markup Language) such as OWL (Ontology Web language) and RDF (Resource Description Framework) languages of ontologies. OM not only describes resources and services like OWL and RDF do. OM represents objects, geographic locations, bibliographies, and even novels (referring to novel phrases, this is not recommended because they are fantastical and unrealistic phrases used that the computer would not be able to interpret).

An example of unrealistic phrase was extracted from the novel: "Cien años de soledad" by Gabriel García Márquez: *"Un gitano corpulento, de barba montaraz y manos de gorrión, que se presentó con el nombre de Melquiades, hizo una truculenta demostración."* This phrase is so complicated to be represented in an ontology because the different senses that could be understood totally different and not according to the words of the sentence. In the case of *"sparrow hands"* the computer would be looking to establish the relationship between the hands of a man and the bird. ANACONJ uses the text better due to the fact that it does not include figurative words.

The OM notation has basic tags like the ones to be described below. For more information we refer to (Cuevas-Rasgado, 2006). The following tags are used to describe ANACONJ semantic frameworks.

1. <**concept**>: this label represents or defines a concept. A concept can be an object as a chemical compound: *Trinitrotoluene*, geographical place: *Querétaro*, personage: *Benito Juárez* (that would be the instance of a set of objects of type person), ideas like "*The city of the hope*", intangible objects like: *Soul*, and emotions like *sadness*, animals: *cat*, plants: *poppy*, transport: *car*, tools: *hammer*, etc. A concept can contain in itself other concepts and thus is the nesting of concepts. The closing tag is </ concept>.

2. <**language**>: it always precedes the <concept> tag. It contains the language that describes the concept; it can be Spanish, English, and French. The fact that OM defines an ontology in French, English, or Spanish implies that all concepts will be related in the same language. The closing tag is </ language>.

3. <**word**>: it is always placed after the <Language> tag. It contains all the synonyms of a concept; for example, if the concept is *person* the <Word> tag would be defining words as *human*, *individual*, and *subject*. Regularly the concepts have at least one synonym. The closing tag is </ word>.

4. <**gloss**>: it always precedes the </ word> tag. It contains the description or gloss of the concept, as its definition appears in the dictionary. The closing tag is </ gloss>.

5. <**subset**>: all concepts have an ancestor except the root concept of the ontology; therefore, the content of the <subset> tag indicates the ancestor of the concept currently in definition.

6. <**relation**> **role**: it represents an explicit or external relationship with which the concept currently being defined is connected, in which case role means external role of the current concept with respect to others. It is possible to realize that concepts with role do not have an agent or a liability.

7. <**relation**> **agent**: it contains those concepts that are active agents, which means the one who performs the action of the verb, for example, in the phrase "*the cat drinks milk*", the concept "cat" performs the action "drinks milk."

8. <**relation**> **passive**: it contains those concepts or passive agents; in the example "the cat drinks milk" it is milk who suffers the action of the verb.

The relations (relation) link an "active agent" concept with a "passive agent" concept, which can be found on the left and right, respectively, of the verb or relationship, in a sentence.

3.3. Coordinating Conjunctions. The most used coordinating conjunctions of Spanish are five:

(1) and (copulative),

(2) *or* (disjunction),

(3) *but*,

(4) *nor*,

(5) *if not* (adversative).

The conjunction **y** transforms into **e** whenever it goes before an **i;** for example, Palestine ***e*** Israel; for its part, the conjunction **o** is transformed into ***u*** whenever it goes before one ***or,*** for example, Defensive or offensive? [12].

Coordinating conjunctions function as nexuses to produce syntactic coordination. These conjunctions can unite linguistic forms that fulfill the same function and possess or do not possess the same category.

ANACONJ identifies the conjunction AND and then makes the links of the semantic tree according to its main function in the sentence, which is considering that it is one of the 8 main functions of Table 1.

Table 1: List of the main functions of conjunction AND.

Num.	Elements to the left and right of the conjunction	Example of the use of the conjunction according to its location in the sentence
1	Subjects	His joy **and** spontaneity enchants him.
2	Verbal Predicates	The prongs hold **and** damage.
3	Attributes	Good friendships are pleasing **and** necessary.
4	Direct complement	They rented house **and** car.
5	Indirect compliment	Sent medicines to Bosnia **and** Senegal
6	Circumstantial complement	They arrived late **and** angry.
7	Subordinate prepositions	Tell them if you are happy **and** if you plan to continue.
8	Non-subordinated sentences	Write stories **and** take care of your children

4. Related Works

4.1. Brief Analysis of the Semantic Analyzers of Texts. There are many works done but most of them are for texts in English; the following are the ones closest to ANACONJ's work (for English language):

The study [13] presents an analysis of wh-complements in Montague Grammar. The difference in sense between that-complement and whether-complements plays an important role in the explanation of the semantic properties of sentences in which they are embedded. We coincide with authors that embedding a complement under a verb semantically corresponds to applying the interpretation of the verb to the sense of the complement, i.e., to a propositional concept.

Verbs such as *know* and *tell* operate on the denotations of their complements, i.e., on propositions, and not on their sense, i.e., propositional concepts.

Different kinds of verbs are analyzed, for example: inquisitive verbs (ask, wonder), verbs of conjecture (guess, estimate), opinion verbs (be certain about), verb of relevance (matter, care), and verbs of dependency (depend on).

Different examples are explained using index; i.e., consider the following arguments, of which one of the premises contains a wh-complement with one or more occurrences of wh-terms such as who, what, and which. For example,

John knows which man walks
Bill walks

John knows that Bill walks

Which proposition is denoted by who walks depends on the actual denotation of walk. If Bill walks, the proposition denoted by who walks should entail that Bill walks; if Peter walks, it should entail that Peter walks. This index dependent character can more generally be described as follows. At an index i, who walks denotes that proposition p, which holds true at an index k if the denotation of walk at k is the same as its denotation at i.

The authors conclude that it is difficult to interpret the denoting proposition just by a question as the only true answer for it. Most of the examples that they have analyzed conclude that there is not a semantic theory on its own field that provides a satisfactory result.

The semantic theory requires a pragmatic theory to work. The authors expect that the semantic theory of wh-complements developed in their article in [13] contributes to the survey of the semantic field.

The authors in [14] identify and validate from a large corpus constraints from conjunctions on a positive or negative semantic interpretation of the conjoined adjectives, a long linear regression models used by these constraints to predict if the conjoined adjectives are the same or different orientations, achieving an 82% of accuracy when the conjunctions are considered independently in this task. Evaluations on real data and simulation experiments show high performance levels: with classifications accuracy more than 90% on adjectives that occur in a modest number of conjunctions in the corpus. The strong point of this article is that decisions on individual words are added to provide information on how to group words into a class and whether to label the class as positive or negative.

In reference to [14], the authors have seen that the mentioned works strive to perform an exhaustive parsing and grammatical analysis but there is a distance between these characteristics and the semantic aspect. In fact the works get very good results but could improve if the semantics were taken into account. FreeLing is a project of several years of maturity and has grown exponentially adding other dictionaries like German and Russian to Spanish, English, and Catalan, but some degrees of error in its representation can be resolved with the use of ANACONJ. The role of ANACONJ is to add semantics to a grammatical parser to enrich it and improve it.

5. Development of ANACONJ

Below is the set of steps to resolve the sense of conjunction in the sentence.

5.1. How the System Works. In the module: *the natural language sentence*, from Figure 6, ANACONJ goes through each word of the sentence entered by the user. For example: María habló y Juan escuchó (*Mary spoke and John listened to her*). After that, in *tagger words* module, a tag is placed on each word (including the punctuation) of the sentence using the EAGLES standard. In the same figure, *Revision of semantics in the sentence* module the core of ANACONJ is done by the *ontology,* which contains nodes and relations.

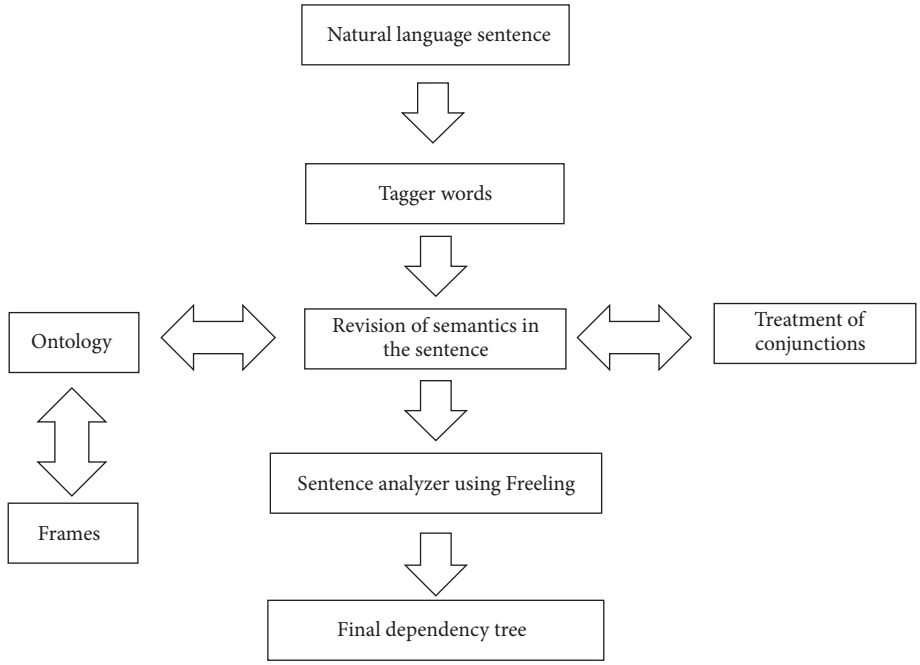


Figure 6: The block diagram of ANACONJ where each block encloses a process and the arrows indicate the sequence of these. The bidirectional arrow indicates that the block sends data and at the same time receives modified data.

We use concepts of common knowledge related by hyponyms and hypernyms, meronyms, synonyms, and semantic frames (*Frames module*) in order to identify the meaning of the conjunction, and here we apply the active and passive agents (seen in Figure 4). The module: *Treatment of the conjunctions* labels the conjunctions as NEWCC and gives the way for using the FreeLing rules, in *sentence analyzer using FreeLing* module, and verifies the partial output of the dependency tree identifying the NEWCC tags. The links are corrected to create the final dependency tree, built up in the *final dependency tree* module. The tree is created based on the corrected links in FreeLing. Figure 6 contains the steps followed by ANACONJ. More details are in [15].

The diagram of *treatment of conjunctions* module is shown in Figure 7.

The way in which ANACONJ works is assigning a weight to the new rules, by relabeling the coordinating conjunctions identified in the sentences; that is, if a coordinating conjunction (y o e) fulfills the function of joining two ideas, this is relabeled as "NEWCC." The new label (NEWCC) does not exist within FreeLing's rules, so no rule applies to conjunctions relabeled as NEWCC.

After relabeling the conjunctions as NEWCC, it is only necessary to assign a very large weight (9999) to the rules that apply to these conjunctions. Because the weight of these rules is very large, FreeLing first resolves the rest of the dependencies of the sentence, to which we apply rules of lesser weight. This is how ANACONJ manages that FreeLing first analyzes the ideas to which the conjunction is joining, and finally both ideas are subordinated to the coordinating conjunction.

Once the new rules have been successfully applied, a correct dependency tree for the sentence is obtained, as shown in Figure 8.

6. Tests

In the case of sentences with conjunctions, we have analyzed sentences in which coordinating conjunctions, specifically the "and" and "e," are combined, joining two different ideas, which we decided to solve. To carry out the tests with this type of sentences the example was taken again: *Maria habló y Juan escuchó (Mary spoke and John listened to her)*.

First, we tested the ideas united by the coordinating conjunction separately, and we identified that FreeLing performs the dependency analysis correctly for both ideas, since in both the verb is the root of the dependency tree. This can be seen in Figures 9 and 10.

Freeling's analysis was followed step by step, which allowed verifying that, in effect, the elements to the left of the conjunction were analyzed as a verbal phrase (*Mary spoke*), and the resulting analysis was the tree of Figure 11.

In this way FreeLing assumes that whenever the conjunction is found and in the sentence, it will serve to enumerate nominal phrases. Under this principle FreeLing will continue his analysis and subordinate the closest noun phrases to the right and left of the conjunction and for the case of this example only finds a noun phrase to the right which is *Juan*.

In order to check what was stated in the previous paragraph, a new test was executed with the sentence "*María habló español y Juan escuchó.*" See Figure 12. Considering FreeLing's rules, the *español and Juan* words should be subordinated to the conjunction *y* (*and*).

As can be seen in Figure 14, FreeLing only considers the conjunction *and* to perform enumerations. Once this process is done, the analysis of the sentence will continue, but it will be wrong to analyze the text to the right of the conjunction, since it will only find a verb (*listened*), since the noun phrase

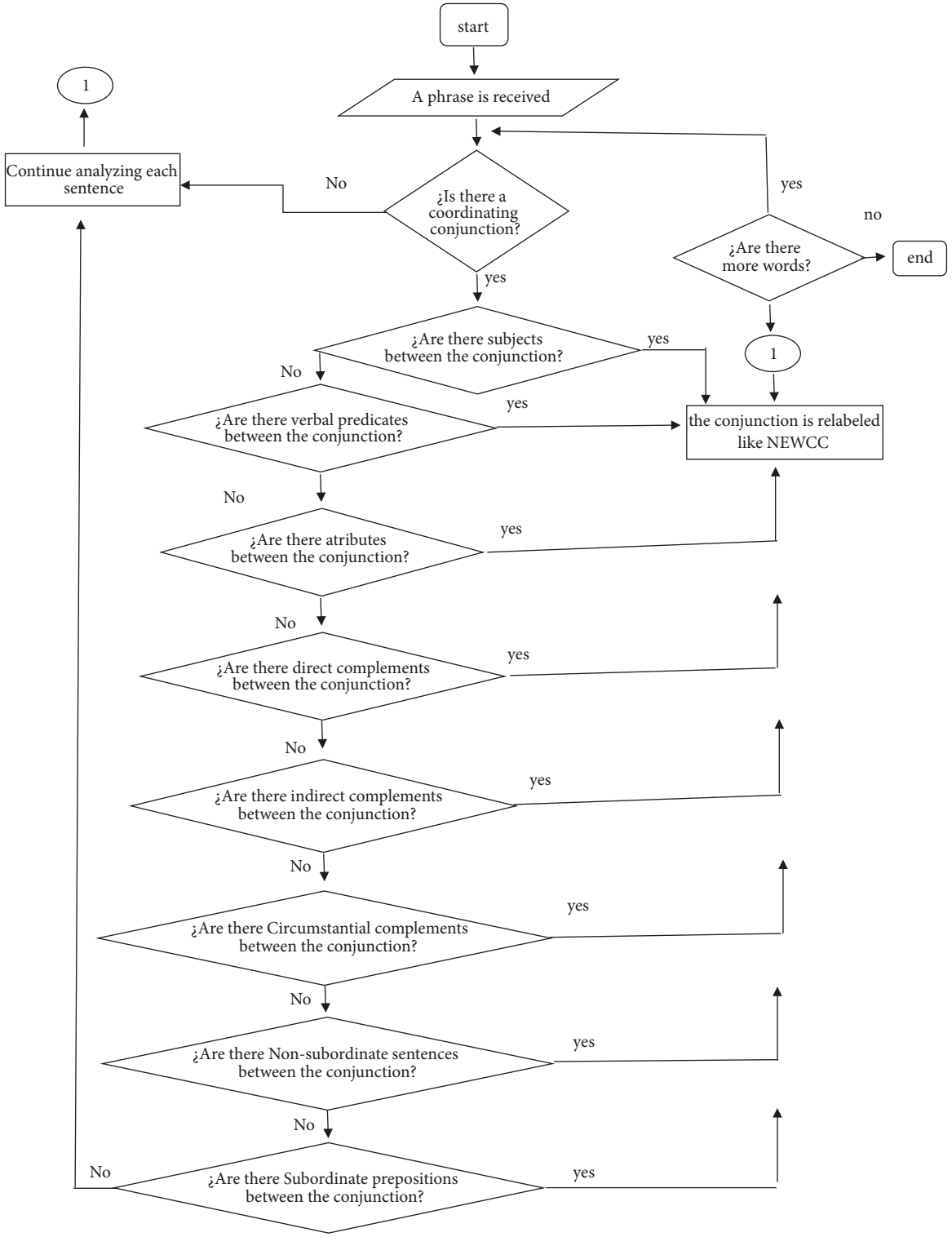


Figure 7: The diagram of the figure details the algorithm for the solution of the conjunctions, and specifically only those conjunctions that are to be labeled are indicated to give a later treatment.

(John) has been subordinated to the same conjunction, and a default rule will apply and subordinate the verb *listened* to the main verb of the sentence (spoken), which is incorrect since one verb cannot subordinate another.

A correct analysis should consider that the conjunction is the core of the sentence since it is linking two different ideas and therefore this should be the root of the dependency tree.

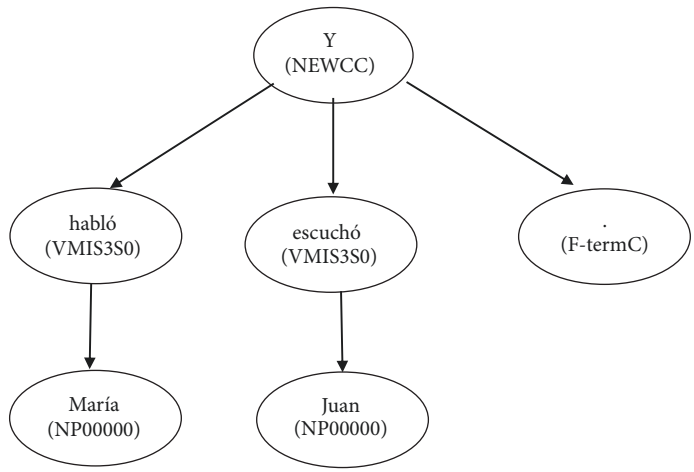


Figure 8: Tree of dependencies according to ANACONJ for the sentence *Maria habló y Juan escuchó* (*Mary spoke and John listened to her*).

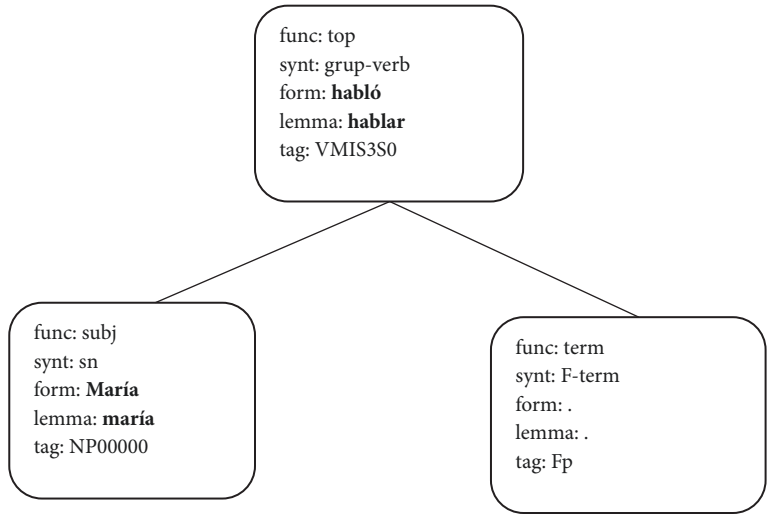


Figure 9: Tree of dependencies for the sentences *María habló (Mary spoke)*.

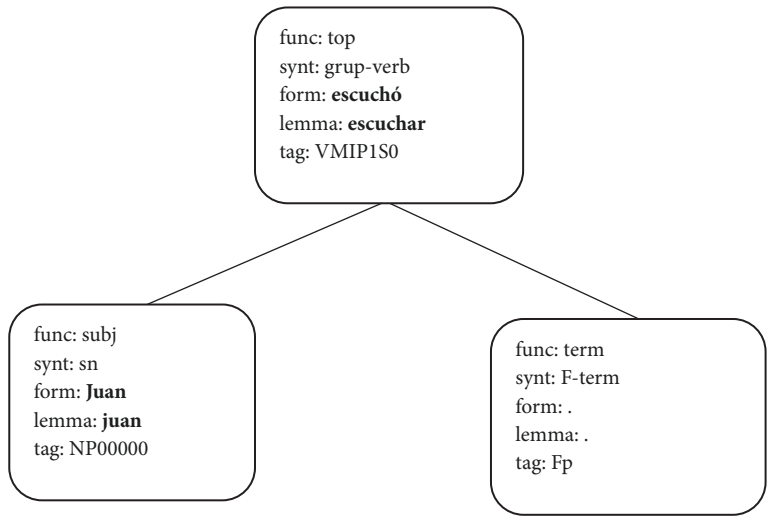


Figure 10: Tree of dependencies for the sentences *Juan escuchó (John listened)*.

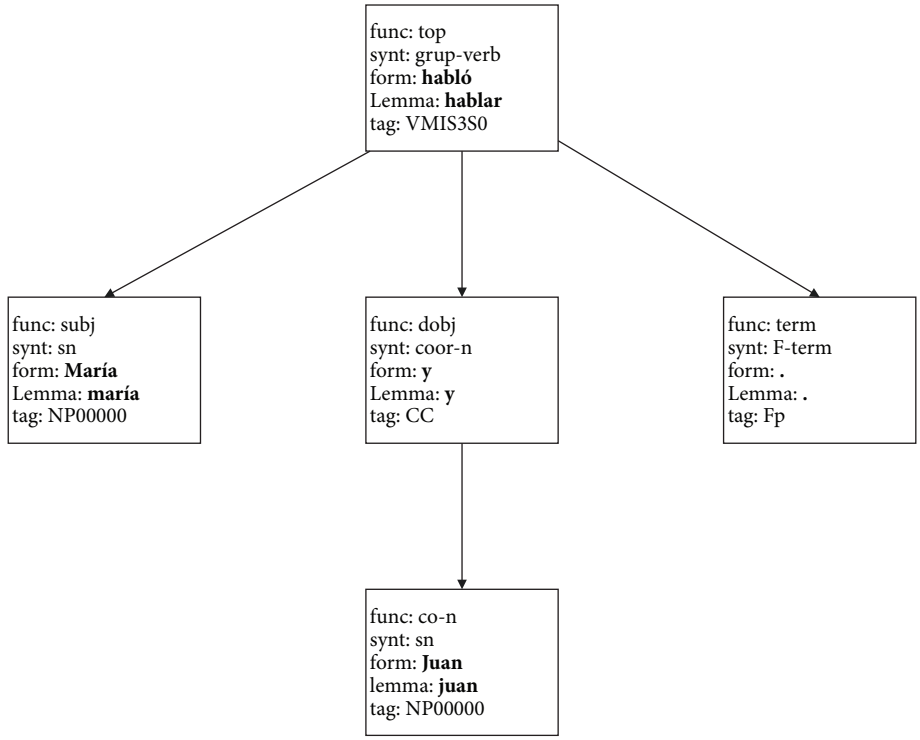


Figure 11: Analysis with FreeLing for the sentence: *María habló y Juan escuchó*.

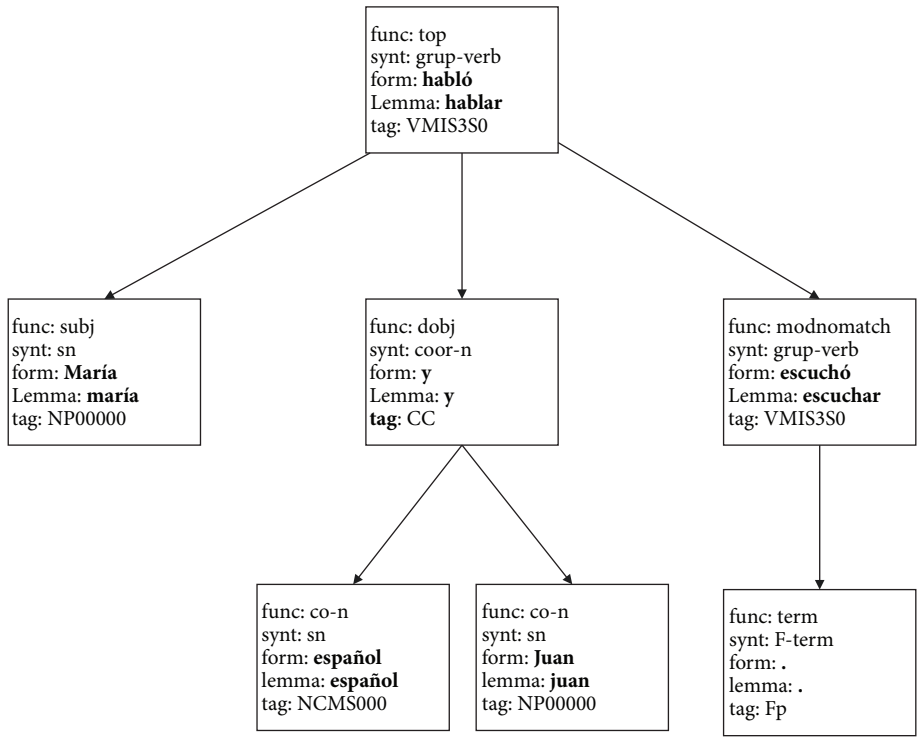


Figure 12: Tree of dependencies for the sentences: *María habló español y Juan escuchó* using FreeLing without ANACONJ.

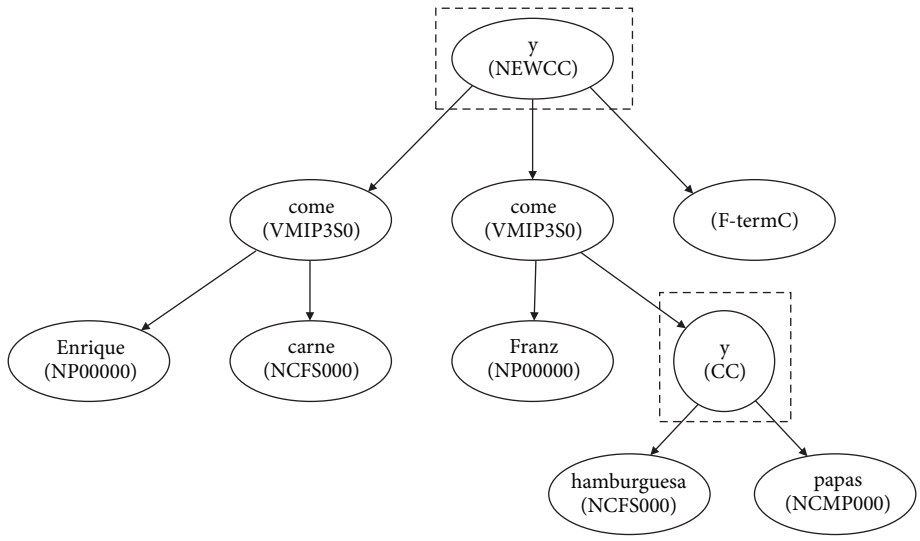


Figure 13: Tree of dependencies according to ANACONJ for the sentences *Enrique come carne y Franz come hamburguesa y papas*.

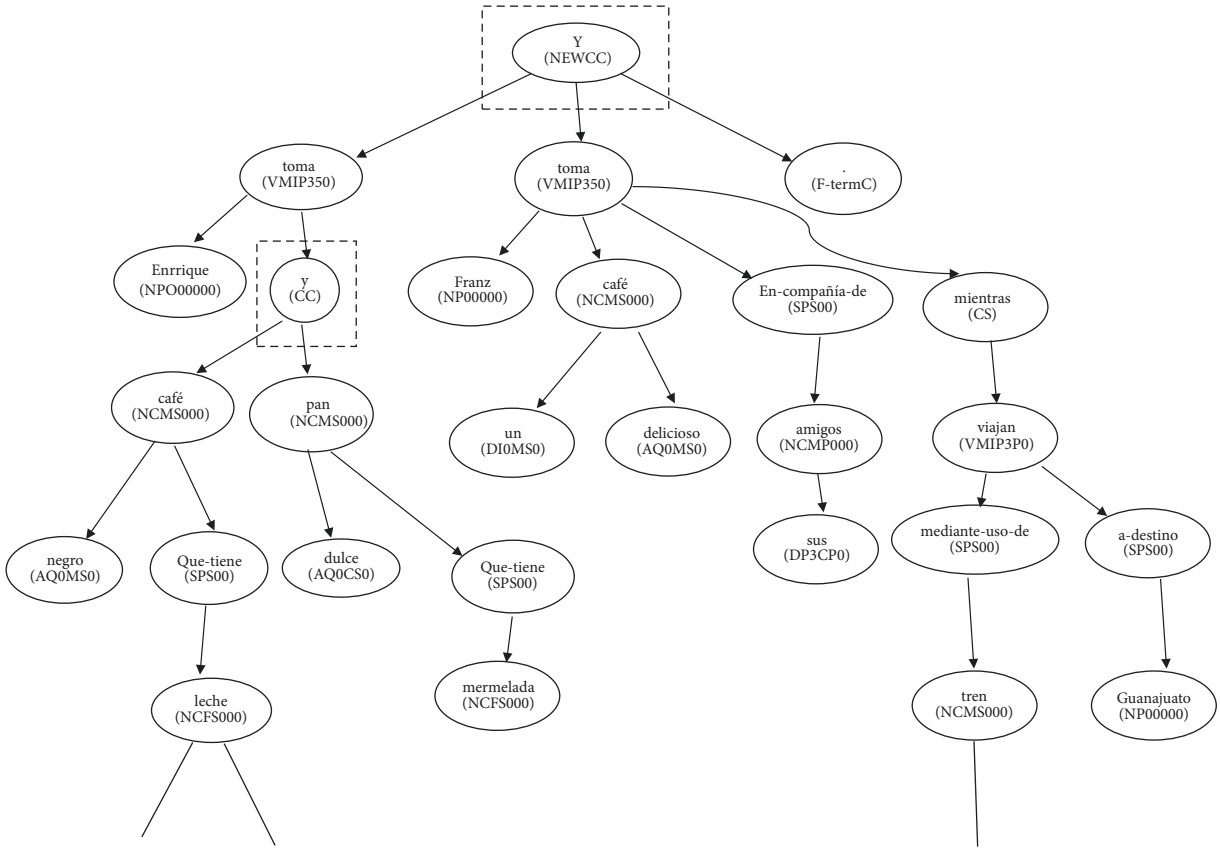


Figure 14: The tree is formed automatically. The generation time of this depends on the number of words in the sentence.

7. Results

We star the tests with two sentences: **first example**, is shown in Figure 13:

Enrique come carne y Franz come hamburguesa y papas, (Enrique eats meat and Franz eats hamburger and potatoes).

Second example: *Enrique toma café negro con deliciosa leche caliente y pan dulce con mermelada y Franz toma un delicioso café con sus amigos mientras viajan en un tren a Guanajuato (Enrique drinks black coffee with delicious hot milk and sweet bread with jam and Franz drinks a delicious coffee with his friends while they travel on a train*

to Guanajuato), The rectangles with dotted line joins verbs, nouns, complements, etc. See Figure 11.

7.1. Examples Using Corpus in Spanish. The **C**onference on **C**omputational **N**atural **L**anguage **L**earning (CoNLL) 2009 (https://ufal.mff.cuni.cz/conll2009-st/task-description.html) corpus is the result of an effort between work teams, for 5 years. These teams have the purpose of promoting applications of Natural Language Processing. The corpus is evaluated under standard configuration, and specifically the Corpus version 2009 is dedicated exclusively to syntactic parsing.

We download CoNLL 2009 Spanish trial, after we tried ANACONJ to identify the conjunction AND, and finally 38 conjunctions were found in 32 sentences; see the next list. For clarity, the conjunctions are **bold char**.

(1) *El gobernante con ganada fama desde que llegó hace 16 meses al poder de explotar al máximo su oratoria* **y** *acusado por sus detractores de incontinencia verbal...*

(2) *. . .aceptó el aplazamiento de los comicios* **y** *valoró la "pedagógica" medida como un triunfo de la democracia venezolana.*

(3) *. . . la gira presidencial prevista entre el 16* **y** *el 25_de_junio por varios países de la Organización_de_Países_Exportadores_de Petróleo (OPEP),..*

(4) *.. la generalidad de las fuerzas políticas venezolanas de ser el verdadero* **y** *único causante del aplazamiento . . .*

(5) *. . . lluvia de críticas por su "falta de experiencia"* **y** *a sus cinco miembros se les ha prohibido la salida del país, . . .*

(6) *. . . el Congresillo quiere guardar las formas* **y** *consultar a la "sociedad civil" . . .*

(7) *. . . sería la composición idónea* **y** *equilibrada del CNE,..*

(8) *. . . deben tener experiencia para "dirigir procesos complejos",* **y** *deben ser personas con independencia* **y** *equilibrio, respetabilidad* **y** *reconocida solvencia moral,...*

(9) *. . .se vio obligado a salir de los ya casi normales* **y** *permanentes rumores de golpe de Estado,* **y** *negó que haya división en las filas castrenses* **y** *mucho menos que esté en marcha una conspiración contra Chávez.*

(10) *. . . los partidarios de Chávez* **y** *los seguidores del aspirante presidencial* **y** *también militar retirado, Francisco_Arias_Cárdenas,...*

(11) *. . .La amnistía favorece a los catorce coroneles detenidos* **y** *al más de un centenar de oficiales . . .*

(12) *. . .Noboa, que fue vicepresidente en el gobierno de Mahuad* **y** *le sucedió en el cargo tras su caída. . .*

(13) *. . . la amnistía permitirá la pacificación de la nación* **y** *la creación de un ambiente propicio para el diálogo* **y** *la concertación...*

(14) ***Y*** *es que los coroneles rebeldes gozan de una amplia simpatía..*

(15) *. . . indígenas que ocuparon el 21_de_enero el Palacio* **y** *luego marcharon hacia el centro . . .*

(16) *. . . la suspensión de procesos penales civiles* **y** *los seguidos en la Corte_de_Justicia_Militar. . .*

(17) *. . .Los frentes populares no están dispuestos a que las fuerzas impongan sanciones* **y** *si, acaso, las aplican, convocarán...*

(18) *"Vamos a estar vigilantes de que se cumpla la Constitución* **y** *que la amnistía sea total...*

(19) *. . . los grupos sociales retomarán las movilizaciones* **y** *se podría llegar "a una convulsión mayor".*

(20) *. . . la amnistía tiene por objetivo "pacificar al país",* **y** *devolver la tranquilidad...*

(21) *. . . otro contenedor amarillo (más pequeño, acoplado al del papel, para reciclar pilas normales);* **y** *otro más pequeño aún (colocado sobre el anterior para las pilas botón).*

(22) *. . .Reciclaje, ahorro, aprovechamiento de los residuos* **y** *lucha contra el despilfarro energético. . .*

(23) *. . . escritor inglés excepcional* **y** *sarcástico proponía . . .*

(24) *. . . reciclarla, guardarla* **y** *volverla a calentar al día siguiente.*

(25) ***Y,*** *en la mesa, se acabó eso de usar los palillos una sola vez* **y** *tirarlos.*

(26) *. . . los restos de comida que hayan quedado adheridos* **y** *ponerlos a secar.*

(27) *. . . cada uno de nosotros evitará que hectáreas* **y** *hectáreas de bosque sean diezmadas...*

(28) ***Y,*** *ya que en eso estamos, . . .*

(29) *ensartarlos en un hilo de nylon,* **y** *alternados con los huesos de las aceitunas . . .*

(30) ***Y*** *nos quemamos.*

(31) *. . . los tres reyes eran blancos* **y** *que en vez de un Niño Jesús,*

(32) *. . . el uno era blanco* **y** *el otro rojizo.*

The CoNLL corpus is not the complete version but it has enough examples that were tested with ANACONJ.

Phrases Found in the Corpus. In the corpus we found 32 phrases that contain 38 instances of conjunction AND that comply with the styles explained in Table 1 (conjunction of 1 to 8). For example, in the sentence number 1: *El gobernante, con ganada fama desde que llegó hace 16 meses al poder de explotar al máximo su* ***oratoria****(NCFS000)* **y** *(NEWCC)* ***acusado*** *(VMP00SM) por sus detractores de incontinencia verbal. . .*

Explanation: in this sentence the conjunction connects to a noun and a verb, being classified according to the sentences of style number 8 of Table 1 (nonsubordinate clauses).

(A) *. . .llegó hace 16 meses al poder de explotar al máximo su oratoria* (. . .he arrived 16 months ago to the power to exploit to the maximum its oratory)

(B) *acusado por sus detractores de incontinencia verbal...* (accused by his detractors of verbal incontinence. . .)

The two events are connected to the conjunction because none depends on the other. That is, one is not a consequence of the other.

Phrases Not Found in the Corpus. Of the 32 sentences in total that contain conjunctions, ANACONJ did not recognize 4 phrases whose structure is not contemplated in its rules; that is, they do not comply with the styles explained in Table 1 and are basically exclamations, for example,

Y *(CC),(FC)* ***en*** *(SP)* ***la*** *(DA0FS0)* ***mesa*** *(NCFS000),* ***se acabó eso de usar los palillos****.* (And, there is not used anymore the toothpicks on the table).

Explanation: in the unrecognized sentence, the conjunction is preserved with the label "CC" and not changed by "NEWCC" as happens when a tree is corrected (see Figure 7) and although it connects to a noun there is no rule (or not was considered in the ANACONJ algorithm) according to Table 1.

Another test was done using the Cast3LB corpus [16], a treebank for Spanish; this treebank is part of **3LB** project that aims to build 100,000 words Spanish. Cast3LB will be enriched in the future with semantic as well as pragmatic information.

We specifically apply ANACONJ to the treebank that contains 2,294 conjunctions AND. ANACONJ did not recognize 926 conjunctions that are not in Table 1 and also those whose frames are not defined in his ontology and those that are not considered in his set of rules.

8. Conclusions

In this work we present the ANACONJ system that recognizes 8 different styles of using AND conjunction according to the semantics of the sentence in which it is found. It works very well for descriptive documents but has been tested with CoNLL09 with good results. ANACONJ had lower precision with Cast3LB corpus because the number of frames of the ontology was less than the number of nouns and verbs of the corpus.

ANACONJ allows the creation of a semantic tree through patterns and rules. According to the achieved results, ANACONJ is able to identify the semantics of more than 90% of conjunctions.

Some specific examples were given in Section 7. The analyzed corpus was obtained from http://ufal.mff.cuni.cz/conll2009-st/. We find that the methods used in ANACONJ seem to work reasonably well for descriptive texts in Spanish. The semantic tree that ANACONJ produces could be used as a reference in concept maps that are very useful at the elementary and high school level. At this educational level the student uses the reading and comprehension of Spanish texts of a topic, identifying their properties, types, and characteristics and differentiating it or relating it to similar concepts. Another use of ANACONJ is in service robots; this robot receives an order and builds its semantic tree finding the most appropriate meaning of the order. For example, the robot could identify a sentence has more than one action to be performed, *ve a la cocina y trae un plátano* (go to the kitchen and bring a banana).

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] D. Villanueva, A.-D. Cuevas-Rasgado, O. Juárez, and A. Guzmán-Arenas, "Using frames to disambiguate prepositions," *Expert Systems with Applications*, vol. 40, no. 2, pp. 598–610, 2013.

[2] F. Colorado, *Mapping words to concepts: disambiguation*, Instituto Politécnico Nacional, D. F., México, 2008 (Spanish), CIC-IPN.

[3] I. Toledo-Gómez, E. Valtierra-Romero, A. Guzmán-Arenas, A. Cuevas-Rasgado, and L. Méndez-Segundo, "AnaPro, tool for identification and resolution of direct anaphora in spanish," *Journal of Applied Research and Technology*, vol. 12, no. 1, pp. 14–40, 2014.

[4] R. Morales, *Resolución automática de la anáfora indirecta en español*, CIC Instituto Politécnico Nacional, México, 2003 (Spanish).

[5] S. Galicia, *Análisis sintáctico conducido por un diccionario de patrones de manejo sintáctico para lenguaje español*, Instituto Politécnico Nacional, D.F., México, 2000.

[6] A. Gelbukh and S. Grigori, *Procesamiento automático del español con enfoque en recursos léxicos grandes. Instituto Politécnico Nacional*, Instituto Politécnico Nacional, D.F., México, 2004.

[7] L. Padró and E. Stanilovsky, "FreeLing 3.0: TowardsWider multilinguality," in *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, pp. 2473–2479, turkey, May 2012.

[8] A. Toral, S. Ferrández, and A. Montoyo, "EAGLES compliant tagset for the morphosyntactic tagging of Esperanto," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP 2005*, pp. 551–555, bgr, September 2005.

[9] M. Minsky, "A framework for representing knowledge," in *MIT-Al Laboratory Memo 306, Reprinted in the psychology of computer vision*, P. Winston, Ed., McGraw-Hill, 1975.

[10] A. Guzmán-Arenas and A.-D. Cuevas, "Knowledge accumulation through automatic merging of ontologies," *Expert Systems with Applications*, vol. 37, no. 3, pp. 1991–2005, 2010.

[11] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, no. 2, pp. 199–220, 1993.

[12] R. Almela, M. Caro, and G. Lozano, *Guía para los exámenes de Lengua castellana y de Comentario de texto*, Ediciones de la Universidad de Murcia, España, 2a edition, 2009.

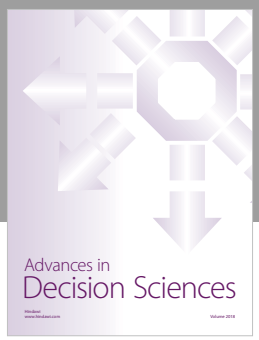
[13] J. Groenendijk and M. Stokhof, "Semantic analysis of wh-complements," *Linguistics and Philosophy*, vol. 5, no. 2, pp. 175–233, 1982.

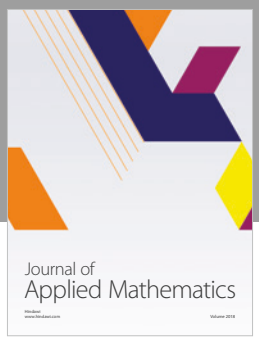
[14] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proceeding of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistic*, pp. 174–181, 1998.

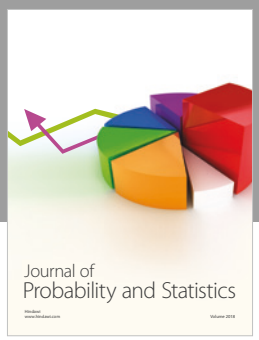
[15] J. B. Arana-Roa, F. L. Lake-Moctezuma, and E. A. Vergara-Albarrán, *Implementing a module for semantic analysis upon*

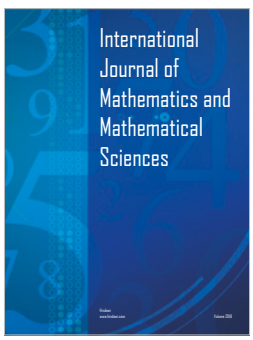
the syntactic analyzer FreeLing, ESCOM Instituto Politécnico Nacional, D. F., México, 2015 (Spanish), No. 2014-A024.

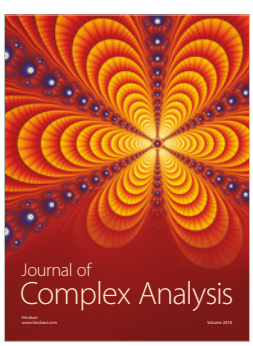
[16] M. Civit and M. A. Martí, "Building Cast3LB: A Spanish Treebank," *Research on Language and Computation, Springer*, vol. 2, no. 4, pp. 549–574, 2004.

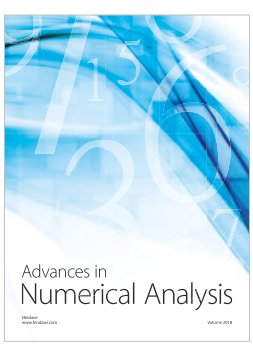


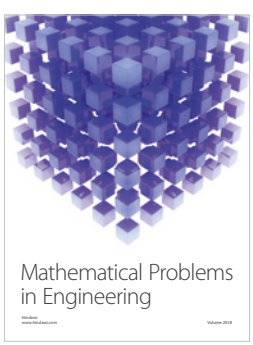


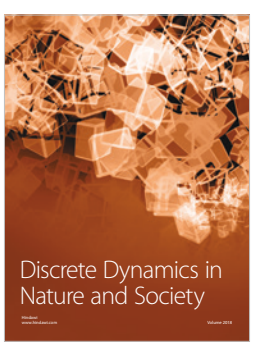


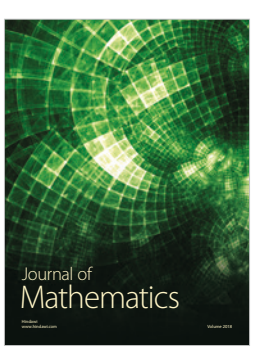


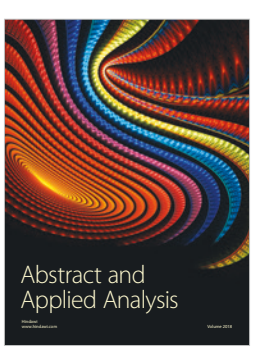


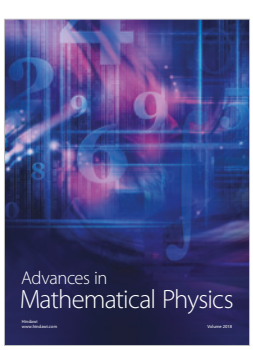












Hindawi

Submit your manuscripts at
www.hindawi.com