

Functional Disambiguation Based on Syntactic Structures

Octavio Santana Suárez, José Rafael Pérez Aguiar,
Luis Losada García, and
Francisco Javier Carreras Riudavets
University of Las Palmas de Gran Canaria

Correspondence:

Francisco Javier Carreras
Riudavets,
Departamento de
Informática y Sistemas,
Edificio de Informática y
Matemáticas, Campus
Universitario de Tafira,
Universidad de Las Palmas
de Gran Canaria, 35017 Las
Palmas de Gran Canaria,
Las Palmas, Spain.

E-mail:

fcarreras@dis.ulpgc.es

Abstract

This article presents a disambiguation method which diminishes the functional combinations of the words of a sentence taking into account the context in which they appear. This process is built in two phases: the first phase is based on the local syntactic structures of the Spanish language and reaches an average yield of 87%. The second one is supported by syntactic tree representation and pushes the results up to an approximate high end of 96%. This process constitutes the starting point towards an automated syntactic analysis.

1 Introduction

In the Spanish language, there are a considerable number of words that can play different grammatical functions, and therefore a text analysis would produce an enormous amount of combinations unless the function of each word within the context where it appears is considered. Functional disambiguation consists of the elimination of the results that do not answer to their function within the text.

This article presents a method of functional disambiguation this method reduces the size of the answer through a two-step treatment of a morphological processor. In the first stage, a functional disambiguation based on local syntactic structures is applied; here the grammatical functions that invalidate the neighbouring environment of every word within the sentence are discarded. In the second stage, the functional disambiguation is performed; at this point the combinations of grammatical functions of the sentence that prevent the generation of syntactic representation trees valid for the whole sentence are discarded.

2 Basic Syntactic Structures and Functional Pairs

In the Spanish language, there are basic structures that repeat and combine over and over among themselves in order to give way to the sentences of the discourse. The composition of these structures defines the pairs of grammatical functions that appear in a sentence—within these local-type structures. When a local-type study is to be performed, the null symbol is included both at the beginning and at the end of every structure.

The functional behaviours of the following need to be considered: noun, adjective, demonstrative adjective, possessive adjective, adverb, personal pronoun, relative pronoun, remaining pronouns, article, preposition, conjunction, coordinating conjunction and contraction.

Some categories are disclosed because they show function and position differences in the syntactic structures. Among adjectives it is possible to distinguish the possessive ones from the demonstrative ones; the possessive adjectives that can appear before, after and in both positions in relation

to the nominal head which they complement can be separated. Among pronouns we can with distinguish the demonstrative adjectives—with adjective function—from the personal pronouns—the unstressed and the tonic pronouns are identified separately—and from the relative pronouns; the remaining group of pronouns are considered under the denomination of other pronouns. The coordinating conjunctions will be taken into account in a special fashion because they are used to link formal structures of the same syntactic level—all of them are included under the denomination of a conjunction. The personal forms of the infinitive, gerund and participle can also be distinguished from each other.

Among the contracted forms, a combination of a preposition and a determiner and, sometimes, a combination of three elements will be considered. The punctuation marks are also considered, differentiating between a comma and a semicolon.

2.1 Homogeneous noun phrase

The homogeneous noun phrase has the following basic structure: *null + determiner + nominal head + adjacencies + null*. The determiner may be: an article, a possessive adjective or a demonstrative adjective. The nominal head is formed by a noun. The adjacency may be: an adjective, the *de* preposition followed by a phrase (prepositional complement of the noun) or a noun (apposition). The determiner, the nominal head and the adjacency must agree in gender and number.

This structure may exhibit certain variations in relation to the presence and position of their elements. The nominal head will always be present; however, the determiner and the adjacency may not appear. Some adjacencies—adjective—may precede the nominal head and, sometimes, the determiner—possessive adjective—may follow the nominal head.

Table 1 shows the configurations formed by consecutive pairs are shown.

2.2 Heterogeneous noun phrase

Heterogeneous noun phrases are combinations of the homogeneous ones: *homogeneous noun phrase + connector + homogeneous noun phrase*.

Table 1 Pairs of symbols that form *homogeneous noun phrase*

Followed by	null	determiner	nominal head	adjacency
null	no	yes	yes	yes
determiner	yes	no	yes	yes
nominal head	yes	yes	no	yes
adjacency	yes	yes	yes	yes

Table 2 Pairs of symbols that form *heterogeneous noun phrase*

	Preceded by connector	Followed by connector
determiner	yes	yes
nominal head	yes	yes
adjacency	yes	yes

The connectors are conjunctions, from the grammatical point of view, and from the graphical point of view they are realized by the comma (,). The new combinations of symbols that appear are listed in Table 2.

2.3 Substitute noun phrase

The substitute noun phrase appears when the nominal head is realized by a category different from the noun: the pronoun, the adjectives and the infinitives preceded by a determiner. With respect to the homogeneous noun phrase, they appear as new pairs of functional categories as a result of substituting the noun that forms the head by a pronoun (tonic personal pronoun, pronoun of relative preceded by an article or other pronoun), an adjective (preceded by an article or a demonstrative adjective) or an infinitive (preceded by an article, possessive adjective or demonstrative adjective). The head continues to conform with the determiner and the adjacencies with regard to gender and number.

2.4 Verb

The simple verbal forms are constituted by one verb in the active mode with the basic structure: *null + simple verbal form + null*. A simple verbal form may be a personal verbal form or an infinitive.

A complex verbal form has the following basic structures: *null + auxiliary + impersonal form + null* and *null + proclitic + personal form + null*. The first structure includes the already understood auxiliary followed by a participle, compound tenses, passive voice, the auxiliary of direct action followed by an infinitive and an auxiliary verb followed by a gerund. The second structure represents an unstressed personal pronoun followed by a simple or a compound personal verbal form.

In addition, there are two special cases that must be treated: *null + indirect incidence auxiliary + conjunction + infinitive + null* (because the only acceptable conjunction is *que*) and *null + indirect incidence auxiliary + preposition + infinitive + null* (because the acceptable prepositions are *a*, *de*, *en* and *por*).

Finally, the existence of multiple verbal heads have to be taken into account: *verb + connector + verb* (the connectors are again conjunctions from the grammatical point of view and the comma from the graphical point of view).

2.5 Prepositional phrase

The prepositional phrase comprises a preposition plus a noun phrase; the pairs of contributing functional categories are combinations with the preposition preceded by *null* or followed by a determiner, a nominal head, an adjacency and another preposition (in the case of double preposition, the first one has to be *a* or *hasta*).

2.6 Adjectival phrase

The simple adjectival phrase, which exists only with copulative verbs, acts as an attribute, and it is formed by an adjective in the basic structure: *null + adjectival phrase head + null*. The adjectival phrase head is an adjective.

In the multiple adjectival phrases, adjectives may appear linked by connectors: *null + adjectival phrase head + connector + adjectival phrase head + null*; the connector is a coordinating conjunction or a comma.

2.7 Adverbial phrase

The syntactic structures of the adverbial phrase are: *null + adverb + null*, *null + adverb + adverb + null*,

null + adverb + prepositional phrase + null, *null + adverb + nominal phrase + null* and *null + adverbial phrase + null*.

The pairs of contributing functional categories are combinations where the adverb is preceded by *null* and a preposition (in the case of an adverbial phrase), or followed by another adverb, a preposition, a determiner, a nominal head or an adjacency.

The adverb may appear, in some cases, adjacent to an adjective within a noun phrase; because the combination: *definite article + adverb* must be added.

2.8 Linkage among several structures

The basic structures are combined in order to generate structures of a larger size. In many cases, it is not necessary to use linking particles, but in others it is. When the structures to be linked are clauses, it is necessary to use a linking element. For this reason the pairs: *null + linking element* and *linking element + null* are added; the linking element can be a conjunction, a comma or a semicolon.

3 Local Functional Disambiguation

Local functional disambiguation starts from the following data:

- (1) The allowed set of functional behaviour S —referred to in Section 2.
- (2) The set of pairs P of symbols of the form $a + b$, where a and b belong to S , that can exist in the local structures in the Spanish language—they have been presented in Section 2.
- (3) A set of combinations of functional categories that are not allowed.

Due to the existence of rules of the form *null + category* and *category + null*—beginnings and endings of local structures—disallowed combinations may occur; to avoid this, a set of functional structures that are not prohibited has to be defined (Table 3).

Table 3 Prohibited combinations

Prohibited functional combinations
preceding possessive adjective + preposition
preceding or after possessive adjective + preceding possessive adjective
definite article + preposition
comma + conjunction + punctuation
conjunction + null
null + demonstrative adjective + personal verbal form
pronoun + infinitive
unstressed personal pronoun + adjective
unstressed personal pronoun + adverb
unstressed personal pronoun + determined article
unstressed personal pronoun + conjunction
unstressed personal pronoun + preposition
unstressed personal pronoun + pronoun
unstressed personal pronoun + noun
punctuation + conjunction + comma
noun + participle
...

(4) Starting from the words that produce the combination, a set of special cases is defined:

- When there is more than one verbal form without a link relation, the option is disregarded.
- If there is a determiner but no adjacencies, the ambiguity between the adjective and the noun is resolved in favour of the noun because it is the head of the noun phrase.
- In the case of ambiguity between adjective and participle, the term favoured is the adjective if there are no auxiliary verbs—*haber* and *ser*.
- In order to avoid problems of cacophony, the concordance of the *de* article with the nominal head which it precedes is not necessary.
- Before *mí*, *ti* or *sí*, only a preposition may appear.
- After a question or exclamation mark, *qué* will be another pronoun.
- After a verb or an adverb, *que* will be a conjunction; after *el*, *la*, *las*, *lo*, *los* it will be a relative pronoun; after a comma it will be a pronoun or a conjunction.

- Before a noun, *de* will be a preposition.
- The word *no* acts only as a noun after *el* or *un*.
- The words *sobre* and *muy* do not have the value of a noun before another noun.

The following steps are executed:

- (1) Processing of the morphological analysis of the sentence and getting a set of potential functional combinations.
- (2) Examining all strings in groups of three elements in order to accept or reject the central element. Given the sequence of functions $a + b + c$, b is accepted if and only if it is given any one of the following conditions:

- (1) $\{a + b\}$ and $\{b + c\}$ belong to P
- (2) $\{null + b\}$ and $\{b + c\}$ belong to P
- (3) $\{a + b\}$ and $\{b + null\}$ belong to P
- (4) $\{null + b\}$ and $\{b + null\}$ belong to P

- (3) Of the sequences not rejected, those containing any prohibited functional elements are eliminated.
- (4) The remaining combinations also include some that fit the special cases.

4 Structural Ambiguities

Starting from a combination of functional behaviours of the words of a sentence, it is possible to get more than one tree for the analysis when applying the Spanish grammar considered here—it is formalized with more than 400 rules; such multiple results denote structural ambiguity.

The existence of more than one rule with the same symbol or combination of symbols on the right side is what is denominated direct structural ambiguity; the grammar used here comprises more than 60 direct structural ambiguities that cover about 240 rules. The direct structural ambiguities led to primary conflicts. There are cases of real ambiguities that can lead to more than one valid interpretation of a sentence.

5 Solving Primary Conflicts

In the following paragraphs several proposals for producing rules to resolve conflicts will be considered; the superposition of these rules will produce the removal of the non-acceptable trees of analysis. In some cases, the rules may be applied the very moment a new symbol is added during the process of analysis, i.e. when the rules depend upon the symbols of the lower levels; in other cases it would be necessary to wait until the completion of the tree.

5.1 Ambiguities and necessary words

For some of the complements it is not possible to use all the words of a given functional behaviour: in this sense any pronoun neither originates a direct object nor does any preposition of a prepositional phrase give way to an indirect object.

Rule: Necessary Words

Let S be a non-terminal symbol generated starting from an intermediate symbol IS and let $PN(S)$ be the set of words necessary for S , the S symbol will be accepted if and only if it is found to belong to the set $PN(S)$ among the words generated by IS .

5.1.1 Prepositional Phrases

Various structures can be generated from IS = prepositional phrase; however, the allowed preposition are not the same for all the structures.

The direct object, for example, only takes *a* and the indirect object takes *a* or *para*. In this sense, conflicts are eliminated in some cases and in the remaining ones the conflicts are diminished.

In the case of concatenating prepositions the same rule is applicable, and is applied to the second

preposition. When a contraction appears, the same considerations for the preposition are to be applied.

There are words that, in general, are not recognized as prepositions, but they have similar functional behaviours; they are the so-called imperfect prepositions.

5.1.2 Unstressed personal pronouns

Various structures can be generated starting from IS = unstressed personal pronoun; however, the allowed pronouns are not the same for all the structures (Table 4).

5.1.3 Other categories

Other categories used in resolving conflicts are shown in Table 5.

5.2 Ambiguities and symbols not allowed

If starting from an unstressed personal pronoun, a substitutive noun phrase is generated. It should not give place to a direct object because such pronouns

Table 4 Unstressed personal pronouns in the resolution of conflicts

S = Structure	NW(S) = Necessary Words
direct object	<i>la/s, lo/s, me, nos, os, se, te</i>
indirect object	<i>la/s, le/s, lo/s, me, nos, os, se, te</i>
attribute	<i>lo</i>
morpheme of passive construction	<i>se</i>
morpheme of impersonal construction	<i>se</i>
...	...

Table 5 Other categories in the resolution of conflicts

S = Structure	IS = Category	NW(S) = Necessary Words
adjacency	adverb	<i>como, más, menos, no, todo/a</i>
adjacency	relative pronoun	<i>cuyo/a, cuyos/as, que</i>
subordinate connector	adverb	<i>apenas, como, conforme, cuanto, donde, mientras, siempre, tal, tan</i>
subordinate connector	conjunction	<i>aunque, con que, cuando, cuantos/as, para, porque, que, si</i>
comparative construction	adverb	<i>así, como</i>
adjectival group	adverb	<i>como</i>
...

Table 6 Non-Allowed Symbols (NAS)

S = Symbol	IS = Intermediate Symbol	NAS(S,IS) = Non-Allowed Symbol
direct object	noun phrase	infinitive
indirect object	noun phrase	unstressed personal pronoun
subject	noun phrase	tonic personal pronoun
noun phrase	nominal head	adjective
...

Table 7 Necessary Symbols (NS(s))

S—New Symbol	NS(S)—Necessary Symbol
attribute	copulative verbal head
copulative verbal head	attribute
passive verbal head	passive auxiliary
direct object	verbal head
attributive sentence	attribute
supplement sentence	supplement
intransitive sentence	verbal head
passive sentence	passive verbal head
transitive sentence	direct object
...	...

may have a direct object function only when they are preceded by a preposition.

Rule: Non-Allowed Symbols

Let S be a non-terminal symbol, generated from an IS symbol and let $NAS(S,IS)$ be a set of symbols generating IS , and catalogued as non-allowed generators of S . Then, S generated by IS will be rejected if IS has been generated by means of some symbol of the set $NAS(S,IS)$ (Table 6).

5.3 Ambiguities and related symbols

Some symbols may not appear without the existence of other symbols in the same tree of analysis.

Rule: Necessary Symbols

The symbol S is added to the tree of analysis only if it exists as the $NS(S)$ symbol (Table 7).

In order to reduce the appearance of direct objects erroneously recognized as such, it is advisable to take into account the need for having a transitive verb. Taking also into account that copulative verbs are of the intransitive type the

possibility of confusing an attribute with a direct object is, hence, reduced.

Rule: Necessary Symbols with Condition

The S symbol is added to the tree of analysis if and only if it exists as an NS symbol that complies with the condition $C(S, NS)$.

5.4 Ambiguities and incompatible symbols

The differentiation between ambiguities and incompatible symbols is based on the non-existence of a symbol and not on its existence; in order to accept the intransitive sentence symbol the direct object symbol must not exist. Every tree that includes incompatible symbols is rejected—with the exception of compound sentences which consist of several predicates.

Rule: Incompatible Symbols

The symbol S is added to the analysis tree if and only if the $InS(S)$ symbol does not exist (Table 8).

5.5 Concordances

Among the different structures which constitute a sentence there are mandatory requirements regarding the concordance of certain characteristics.

Rule: Concordances

If $S1$ and $S2$ are the symbols of an analysis tree, this tree is accepted if and only if there is concordance between the set of definite characteristics for these symbols, $CSD(S1, S2)$.

5.5.1 Description of cases

Concordances that have been checked during the process of local functional disambiguation should be verified again because two elements that appear in the same phrase must concord and in

Table 8 Incompatible symbols

S—New Symbol	InS(S)—Incompatible Symbol
attribute	attribute
attribute	morpheme of passive
attribute	direct object
attribute	indirect object
attribute	objective predicative
attribute	subjective predicative
attribute	supplement
agent complement	supplement
morpheme of impersonal	morpheme of impersonal
morpheme of impersonal	morpheme of passive
morpheme of impersonal	morpheme of half voice
morpheme of passive	morpheme of passive
morpheme of passive	morpheme of half voice
morpheme of half voice	morpheme of half voice
verbal head	verbal head
direct object	attribute
direct object	morpheme of passive
indirect object	attribute
indirect object	morpheme of passive
attributive sentence	passive verbal head
supplement sentence	passive verbal head
supplement sentence	direct object
intransitive sentence	attribute
intransitive sentence	direct object
intransitive sentence	supplement
transitive sentence	attribute
transitive sentence	passive verbal head
objective predicative	attribute
objective predicative	objective predicative
objective predicative	subjective predicative
subjective predicative	attribute
subjective predicative	morpheme of impersonal
subjective predicative	subjective predicative
subject	morpheme of impersonal
supplement	attribute
supplement	agent complement
supplement	supplement
...	...

the local analysis it might be that this was not the case because the union of local structures was assumed.

Concordance between subject and verbal head: in sentences with verbs in a personal form there must be concordance in number and person with the head of the subject structure.

$$CSD(\text{subject}, \text{verbal head}) = \{\text{number}, \text{person}\}$$

$$CSD(\text{subject}, \text{passive verbal head}) = \{\text{number}, \text{person}\}$$

$$CSD(\text{subject}, \text{copulative verbal head}) = \{\text{number}, \text{person}\}$$

Besides the concordance between subject and predicate, the concordance between the next descendants of the predicate must be accomplished—always in gender and number:

Adjacency with nominal head.

Direct object with direct object.

Indirect object with indirect object.

Objective predicative with direct object.

Subjective predicative with verbal head.

Subjective predicative with subject.

Determiner with nominal head.

5.6 Semantic information

The analysis of the semantic content of the words leads to the elimination of ambiguities—starting from the information given through ideological dictionaries: to generate the symbol circumstantial complement of tense among the words that form it, there must be something that gives information on tense or moment.

Rule: Necessary Semantic

If WS is a set of words that are joined to build the S symbol and if $IMS(S)$ is the set of ideological meanings associated with the S symbol, then S is rejected if there is no word in WS such that the ideological analysis belongs to $IMS(S)$.

To improve the efficiency of the automation of the disambiguation process, it should be easy to create a disposition containing the words with the necessary semantics for all the symbols.

To avoid taking into account words with semantics that must not directly intervene with the symbol, the set of WS words will be formed only with words at the highest level of the representation tree for the symbol—if such a level contains only irrelevant words, determiners, connectors and prepositions, the action is to go to lower levels until relevant words are found.

5.6.1 Ideological relationships and symbols

There are words whose semantic content prohibits the generation of a given symbol from another: a homogeneous noun phrase cannot generate a

direct object if the head is a person—ideological information.

Rule: Incompatible Semantic

If *WS* is a set of words that are joined to form the *S* symbol starting from the *IS* symbol and if *SIMR(S, IS)* is the set of ideological meanings that produce the rejection of the *S* symbol, generated from *IS*, then *S* is rejected if there is some word exists in *WS* such that its ideological analysis belongs to *SIMR(S, IS)*.

5.6.2 Ideological relationships among symbols

Symbols can be used as relationships of the ideological type between the subject head and the verbal head: if the verbal head implies an action and is in active form, the subject should constitute a living being.

Rule: Ideological Relations among Symbols

If *S1* and *S2* are symbols of an analysis tree, this tree is accepted if and only if the ideological concordance is satisfied in the set of definite characteristics for these symbols, *IRS(S1, S2)*.

5.7 Special cases

Special circumstances are applied when the previous methods do not solve the problem.

5.7.1 Clauses

The clause, whether coordinate or subordinate, must have a connector, either a subordinator or a coordinator, inserted before or after the clause. Similarly, any type of sentence that allows any of these clauses should comply with the same conditions. The main clause and the subordinate clause are different in the way they are joined to the remainder of the sentence. The subordinate clause of an infinitive should have an infinitive as the verbal head.

5.7.2 Interrogative sentences and exclamatory sentences

The interrogative and exclamatory sentences are differentiated by the punctuation marks that delimit them.

5.7.3 Double direct object

A double direct object-left dislocation is easily recognizable by the following characteristics: (1) the two elements are found together, (2) the first one is found at the beginning of the sentence, (3) the second one is a pronominal clitic and (4) there must be concordance with regard to gender and number between the two corresponding heads.

Rule: Double Direct Object

If *S* is a root symbol that covers the whole sentence, and if two direct object symbols appear, then *S* is accepted if and only if the direct objects are adjacent, are followed by a verbal head and the second direct object is realized by an unstressed personal pronoun.

It must be taken into account that in compound sentences there may be two direct objects for each personal verbal form.

5.7.4 Elimination of options according to the position of the determiners

Only the following symbols: demonstrative adjective, definite article and other pronoun appear before the nominal head. The possessive adjectives can be divided into those that precede the nominal head—*mi, mis, tu, tus, su* and *sus*—and those that come after the nominal head—*mío, mía, míos, mías, tuyo, tuya, tuyos, tuyas, suyo, suya, suyos* and *suyas*—and those that can appear both before as well as after the nominal head—*nuestro, nuestra, nuestros, nuestras, vuestro, vuestra, vuestros* and *vuestras*.

Rule: Post-head Determiners

If *S* is a symbol that belongs to the group of noun phrases and is generated starting from a sequence of symbols where the nominal *head + determiner* sequence appears, *S* will be accepted if and only if the determiner symbol is found in the *PSS (Post-head Symbol Set)* group of terminal symbols that can follow the nominal head.

Rule: Pre-head Determiners

If *S* is the symbol for the adjective phrase that is generated starting from an adjective symbol, *S* will not be accepted if it is found after a determiner symbol from the *BSS (Before Symbol Set)* group of the terminal symbols that cannot follow the nominal head.

5.7.5 Connectors

There are a number of combinations of words that give place to conjunctive conjunctions: *a consecuencia, a distinción de, a fin de, a fin de que, a lo que parece, a medida que, a menos que, a pesar, a pesar de, ahora bien, ahora que, al menos, al objeto de, al objeto de que, al parecer, al paso que, antes bien, así como, así es que, así pues así y todo, aún cuando*, etc.

5.7.6 Other cases

There are situations in which ambiguities can be resolved starting from considerations regarding the words, grammatical categories and intervening objects.

6 Resolutions of Other Conflicts

There are rules that, without being directly applied to a given primary conflict, serve to eliminate ambiguities.

6.1 Symbols that cannot cover the whole sentence

The structure of a sentence should include a subject and a predicate, or possibly only a predicate. The object of analysis cannot comprise the subject symbol by itself.

The main clause and the subordinate clause are symbols that have been defined to generate the analysis of compound sentences; it is for this reason that a main clause symbol has the same structure as a sentence symbol so that the sentence symbols need not cover the whole sentence.

Rule: Total Symbols

The S non-terminal symbol that covers all sequences to be analyzed is accepted if and only if it is found among the symbols of the set *TSS (Total Symbols Set)* that have allowed the covering of whole sentences.

6.2 Verbal periphrasis

The verbal periphrasis formed by more than two elements generates a complex verbal form only in specific cases: *acabar de + infinitive, deber de + infinitive, dejar de + infinitive, echarse a + infinitive, empezar a + infinitive, estar para + infinitive, explotar a + infinitive, haber*

de + infinitive, haber que + infinitive, ir a + infinitive, llegar a + infinitive, ponerse a + infinitive, romper a + infinitive, tener que + infinitive, venir a + infinitive, volver a + infinitive, etc.

6.3 Considerations regarding predicate symbol generation

The rules that define the structures of the predicate are given through the combinations of elements that can appear in it. In a formal definition of structural type it would be necessary to indicate all the possible combinations; because the placement of the majority of the elements is free, the number of possible structures for the predicate would be enormous; therefore it has been decided to permit all combinations and to prohibit those that are not possible—Table 8 shows pairs of incompatible symbols in the same predicate.

The generation of a predicate symbol should be rejected either when some of its ends are not a beginning or an ending of the generated symbol or when there is an adjacency punctuation mark; it would not be a rejection in the case of subordinate sentences—the existence of subordinate elements would be verified as would the existence of multiple verbal forms.

6.4 Other cases

There are specific situations in which ambiguities can be resolved starting from considerations of words, grammatical categories and intervening objects.

7 Experimental Results

We analysed 776 selected sentences covering the broadest spectrum of casuistry inherent to Spanish grammar. The reliability measure for the disambiguation is given by:

$$G = (p \times 100)/(n - 1)$$

where *p* is the total number of functional combinations minus the number of functional combinations

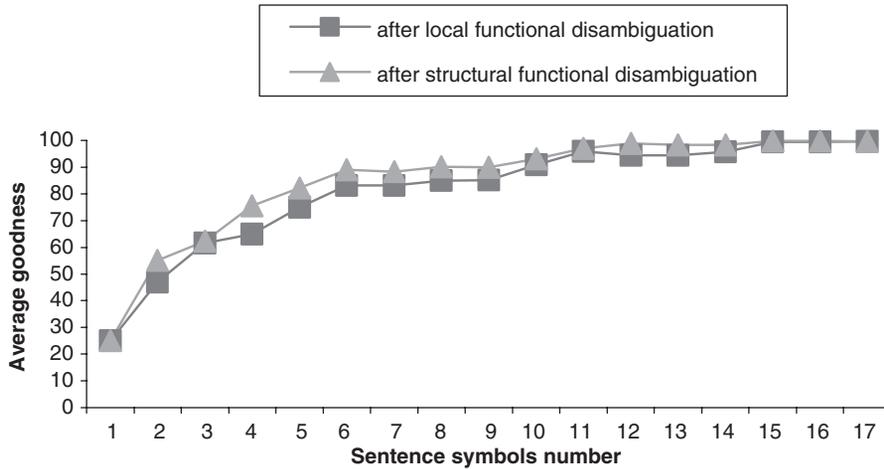


Fig. 1 Functional disambiguation goodness.

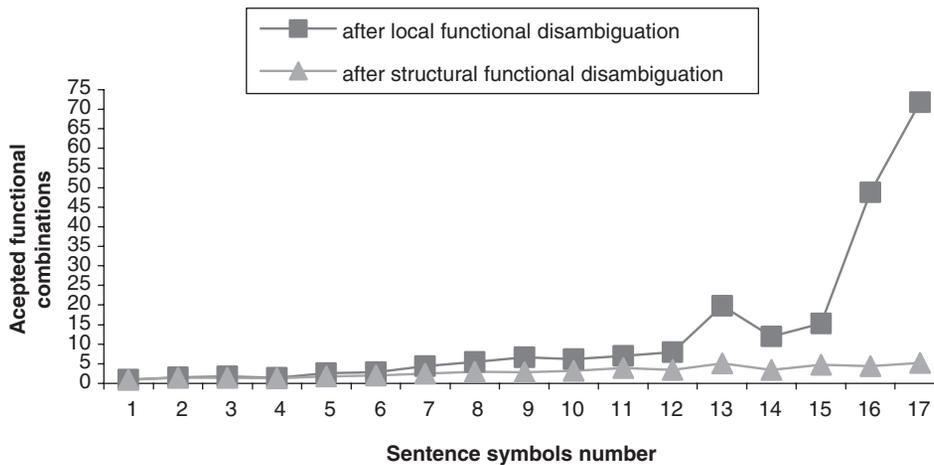


Fig. 2 Number of accepted functional combinations.

accepted and n is the total number of functional combinations provided by the morphological analyser.

As can be seen, in Fig. 1 and Fig. 2, the yield of the functional disambiguation—local and structural—increases with the number of symbols of a sentence. The functional disambiguation based on local syntactic structures has an average yield of 87% and increases to a high of 96% after applying the structural conditions.

8 Conclusions

This study does not stop in subsets of grammar but challenges a whole system of rules for Spanish grammar, despite the notable amount of combinations needed for the analysis.

It contributes towards a solution to the problem of the emergence of functional ambiguities. First, a process of disambiguation based on local syntactic structures is applied; it reaches an average yield of

87%. Subsequently, a disambiguation based on syntactic representation trees is applied. It improves the yield or performance up to a high end of 96%.

The importance of this work lies in its significant contribution to the development of future applications:

- (1) It accelerates the process of syntactic analysis by trimming incorrect structures.
- (2) It improves the precision of results in the advanced searching of words.
- (3) It allows discarding options not valid in the information extraction process.
- (4) It detects grammatical errors in written constructs, etc.

References

- Bosque, I., Demonte, V., and Lázaro Carreter, F. (1999). *Gramática descriptiva de la lengua española*. Madrid: Espasa.
- Gili Gaya, S. (1998). *Curso Superior de Sintaxis Española*. Barcelona: Biblograf S.A.
- Gómez Torrego, L. (2002). *Análisis sintáctico*. Teoría y práctica. S.M., Madrid.
- Quesada, J. F. (1996). Un modelo robusto y eficiente para el análisis sintáctico de lenguajes naturales mediante árboles múltiples virtuales. Centro Informático Científico de Andalucía (CICA).
- Real Academia Española (1989). *Esbozo de una nueva gramática de la lengua española*. Madrid: Espasa-Calpe.
- Santana, O., Pérez, J., Carreras, F., Duque, J., Hernández, Z., and Rodríguez, G. (1999). FLANOM: Flexionador y lematizador automático de formas nominales. *Lingüística Española Actual XXI*, 2: 253–97.
- Santana, O., Pérez, J., Hernández, Z., Carreras, F., and Rodríguez, G. (1997). FLAVER: Flexionador y lematizador automático de formas verbales. *Lingüística Española Actual XIX*, 2: 229–82.
- Santana, O., Pérez, J., Losada, L., and Carreras, F. (2002). Hacia la desambiguación funcional automática en Español. *Procesamiento del Lenguaje Natural*, 28(SEPLN): 1–22.