Semantic-driven Algorithm for Definite Description Resolution

Rafael Muñoz and **Manuel Palomar**

Departamento de Lenguajes y Sistemas Informáticos
Universidad de Alicante
Alicante, Spain
{rafael,mpalomar}@dlsi.ua.es

Abstract

In this paper a semantic-driven algorithm for Spanish definite descriptions resolution is presented. The resolution procedure consists of two main steps: a) the generation of a semantic network using the Spanish WordNet ontology in order to distinguish between anaphoric and non-anaphoric definite description, and b) the application of a set of constraints and preferences to candidates in order to obtain the correct antecedent of an anaphoric definite description. Experiments show a precision of 88.8% and a recall of 83.9% for definite description identification (anaphoric or not) and resolution.

1 Introduction

We commonly use different expressions to refer to a person, an object, an event, a place or a process. These expressions usually used in this way are pronouns and definite descriptions. Definite descriptions (DD) are more difficult to treat than pronouns because the DD do not always refer to an antecedent but may introduce a new entity in the discourse. Furthermore, when a DD is detected we do not know if it is an anaphoric or a non-anaphoric expression. Moreover, the distance between the DD and its antecedent can be greater than the distance between the pronoun and its antecedent because DD have more semantic information than pronouns. For this reason, we need to store all previous noun phrases (NP) in the text. This has the disadvantage of making the system very slow for large bodies of text. This is due to the fact that we need to compare every previous NP with the DD in order to select the correct antecedent. We only consider as antecedent NPs, and reject other kind of possible antecedents like paragraphs or full text.

In this paper we propose a resolution algorithm for Spanish definite descriptions based on semantics. The algorithm automatically generates a semantic network in order to classify DD as anaphoric or non-anaphoric before applying the resolution algorithm. The construction of the semantic network is based on the extraction of the DD ontological concept from the Spanish Wordnet ontology. The algorithm only treats identity coreference without taking in consideration other kinds of coreference such as set-member or part-whole. Moreover, the algorithm solves both direct anaphora (produced by DD with the same head noun as their antecedent) and indirect anaphora (produced by DD with different head noun as their antecedent).

2 The Algorithm

The algorithm contains the following three main components.

- Semantic Network generation module that uses Spanish WordNet to help to identify anaphoric and non-anaphoric DD.
- A set of semantic constraints that rule out anaphoric DD-NP dependence.
- A set of preferences obtained from a empirical study. For each preference a set of values are assigned to several salience parameters (frequency of mention, proximity, semantic relations) for a given NP.

2.1 Generation of Semantic Network

The AGSN (*Automatic Generation of Semantic Network*) procedure builds a semantic network using the ontological concept of head nouns. This ontological concept is extracted from the Spanish Wordnet ontology. The idea is based on the clustering technique although it uses the ontological concept of the head noun instead of a distance. A DD can be referential if there are more than one noun phrase belonging to the same ontological concept because noun phrases are semantically compatible. This mechanism is used to establish a previous identification of anaphoric or non-anaphoric DD. The idea of this mechanism is based on the application of semantic constraints (ontological concept) before running the resolution algorithm, to identify non-anaphoric DD and to reduce the number of candidates for anaphoric DD's. If the DD is the first noun phrase found in the text that belonged in a specific ontological concept, then it is classified as non-anaphoric. But, if it is not the first one then the resolution algorithm is applied. This resolution algorithm only manages as candidates the previous noun phrases in the text that belong in the same ontological concept as the DD. For this reason the number of comparisons is reduced.

The mechanism is based on a simple idea: a DD will be non-anaphoric if it is not semantically compatible to a previous NP. But, if there is one that belongs to the same semantic category, then it can not be classified (*anaphoric* or *non-anaphoric*) without applying the resolution algorithm. The process to generate the semantic network is as follows: the mechanism goes through the text searching for NPs. For every NP, the mechanism extracts its head noun and uses it to look for their ontological concept at the Spanish Wordnet's ontology. If the NP is not a DD, then the NP is added to the list of their ontological concept at the semantic network. If the ontological concept of the semantic network does not exist, the list of this ontological concept is made and the NP is added. However, if the NP is a DD, then the algorithm distinguishes between two possibilities:

The ontological concept does not exist in the semantic network. The ontological concept and the DD are added to the semantic network and it is classified as a *non-anaphoric* DD.

The ontological concept already exists in the semantic network. If there is not a head noun in the same ontological concept semantically related to the head noun of the DD, then we classify the DD as non-anaphoric. But, if it exists any semantic relation between head nouns, the mechanism can not classify directly the DD as anaphoric or non-anaphoric. In this case, an algorithm to solve references is applied to provide the correct antecedent or to classify the DD as non-anaphoric if one is not found. The system can not make a previous classification because two NPs with the same head are not always coreferent (left hand - right hand).

A Word Sense Desambiguation module is needed in order to provide the correct sense of head nouns. The Specification Marks module (Montoyo & Palomar 01) is used in this work to provide the correct sense. After applying this module, a manual review has been made to supervise the results.

2.2 Semantic constraints

After generating the semantic network, a set of semantic constraints for ruling out anaphoric dependence of a DD on an NP due to non-compatible semantic relations is applied. We present two semantic constraints:

R1. Two NPs that belong to the same ontological concept can only be coreferent if they have the same head noun or there is a synonym (car - auto) or hyperonym/hyponym (*hyperonym*: car - vehicle, *hyponym*: car - ambulance) relation between both head nouns.

R2. A comparison between the modifiers of the DD and the NPs semantically compatible is made. If there is an antonym relationship (left - right) between two modifiers then the NP is rejected.

2.3 Preferences

The next step in the process is to apply a set of preferences obtained from a empirical study. For each preference, a set of values is assigned to several salience parameters (frequency of mention, proximity, semantic relations) for an NP. The set of preferences is defined as follows:

P1. Repetition. The system selects previous appearance of the same DD as antecedent (same head noun and same modifiers).

P2. Pre and post-modifiers relation. The system selects candidates with the same head noun and with a semantic relation (synonym, hyperony, hyponym) between pre or post-modifiers of DD and antecedent.

P3. Indirect anaphora (Bridging references[1]). The system selects candidates whose head nouns are related to the head noun of DD through a synonym, hypernym, hyponym relation. Moreover, the system selects these antecedents with pre or post modifiers semantically related.

P4. Antecedent without modifiers. The system first selects candidates with the NP with the same head noun and later with a semantic relationship between head nouns.

P5. Gender and number agreement. The system selects candidates with gender and number agreement.

P6. Closest. If more than one candidate have the same higher salience value, then the system selects the closest candidate. This rule guarantees that only one antecedent is proposed.

2.3.1 Preference Management

There are two different approaches to manage the preference set: an ordered application and a weight management. The ordered application consists of discarding antecedents not fulfilling a preference if there are any candidates that fulfill the preference. The weight management is based on assigning a value or salience to each preference and selecting the candidate with the maximum value.

[1]Definite descriptions with different head noun as their antecedent were called bridging references by Clark (Clark 77)

In our proposal, we use a weight management because it provides the best result (section 3) as also shown in (Mitkov 97). The system works out a salience value for each possible antecedent choosing the candidate with highest salience value as the correct antecedent. Different weights are applied: if the candidate has the same head noun as the DD, then a value of 50 is added to the salience value of the candidate. But, if there is a semantic relation (synonym, hypernym or hyponym) between candidate and DD head nouns, then a value of 40 (synonym relation) or a value of 35 (hypernym/hyponym relation) is added to the salience value. Moreover, an amount for each pair of modifiers of both noun phrases (candidate and DD) is also added to the salience value. This added value is based on the kind of relationship between modifiers. If a modifier appears in both noun phrases then a value of 10 is added to the salience of the candidate. But, if there is a semantic relation between both modifiers, a value of 9 (synonym relation) or 8 (hypernym/hyponym relation) is added to the salience of the candidate. If more than one candidate have the same maximum value then gender and number criteria are applied. Finally, if more than one satisfies the gender and number agreement, a closest preference is applied to guarantee that only one candidate is proposed as antecedent. The salience values applied have been chosen experimentally.

2.4 The Resolution Procedure

The procedure for identifying antecedents of definite descriptions is as follows.

1. Each noun phrase in the current sentence is classified according to its type (definite noun phrase or indefinite noun phrase).

2. The head noun of each indefinite noun phrase is extracted and a list with only one element (indefinite noun phrase) is generated if it did not exist. This list will store the followings noun phrases belonging to the same ontological concept at the semantic network. Otherwise, the NP is stored in the list of the right ontological

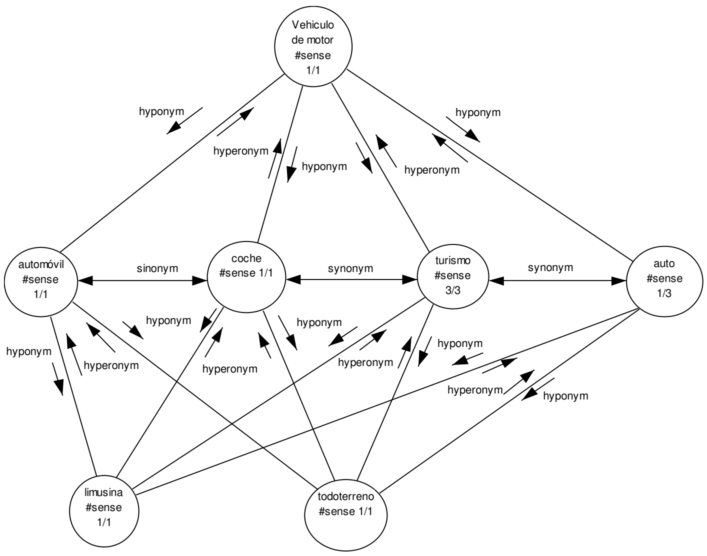

Figure 1: *NPs belonging to the same ontological concept at semantic network*

concept. Figure 1 shows a possible ontological concept at the semantic network made up a set of noun phrases related by semantic relations (synonym, hyperonym and hyponym)

3. For each definite NP, its head noun is extracted and it is checked against all elements of the list of their ontological concept. If the node does not exist then the definite NP is classified as non-anaphoric and the list is generated storing this definite NP. Otherwise, the NP is not classified and the resolution algorithm is applied (step 4).

4. For each NP not classified, a set of semantic constraints to discard antecedents is applied. If all candidates are rejected then the definite NP (DD) is classified as non-anaphoric and it is stored in the list of the ontological concept.

5. For each possible antecedent not rejected, a set of semantic preferences is applied. The preferences are applied only if more than one candidate (possible antecedent) fulfills all restrictions.

6. The NP with the highest value is selected as the correct antecedent.

The algorithm does not limit the number of previous sentences to be used in order to search the correct antecedent for a definite description, i.e., the algorithm stores all previous NPs. However, it only searches the antecedent in the same ontological concept of the semantic network, thus reducing the number of comparisons.

3 Experimental work

The experimentation data was taken from three different corpus: The Lexesp Corpus (L), Deed (notarial) Corpus (D) and Digital Newspaper Corpus (N). The Lexesp[2] corpus is made up of a set of fragments with different kinds of styles (narratives, newspaper articles, etc.) written by different authors. The Deed corpus is a corpus made up of several legal documents used by the information extraction system EXIT (Llopis *et al.* 98). Newspaper corpus is made up of several articles extracted from digital newspapers.

3.1 Training phase

For the training phase, a 43% of the corpus was used. In this phase, several experiments

[2]LEXESP is a Spanish corpus. This corpus has about 5 million of tagged words developed by Psychology Department from University of Oviedo, Computational Linguistic Group of University of Barcelona and Language Treatment Group from Technical University of Cataluña.

were carried out using the training corpora. These experiments involved changes in the constraint and preference set in order to define the optimum configuration[3] to obtain the maximum precision.

For these experiments, the system was trained in order to obtain the best set of preferences. After that, we applied two approaches of preference management to compare both approaches. Once we obtained the best set of preferences and its best management, we evaluated the system with the test corpora to obtain independent results.

The overall results of this version 1 (ordered preferences application) solving the references produced by DD are a precision of 73.2% and a recall of 69.2%. The overall results of version 2 (weight preferences management) solving the references of DD are a precision of 81.5% and a recall of 77%. The weight preferences system increased the score for DD resolution, but it obviously achieved the same scores for the new discourse identification because it uses the same set of constraints. The weight preference management was chosen as the algorithm to be evaluated.

3.2 Evaluation phase

The algorithm was applied to the test corpus (57%) obtaining the results showed in the tabla 1. The systems recognized 759 of 804 non-anaphoric DD. The new entity identification (non-anaphoric DD) is made in two stages. The first stage is made before applying the resolution algorithm. This previous identification classified 304 of 759 DD, similar to non-anaphoric resolution using only the ontological concept (a precision of 40%). The system applies the automatic generation of the semantic network to identify the new discourse entity. Finally, in the second stage, the system classifies as non-anaphoric 406 of 455 (759-304) DD after applying the resolution algorithm. Table 1 shows the performance achieved by the system on the identification task, a recall of 88.3% and a precision of 93.5%.

[3]The term configuration is used to define the set of constraints and preferences that makes up the system for a concrete instance of the experiment process.

Second stage of identification task is developed together with the search of antecedent (anaphora resolution task). We can distinguish between two different sub-task: direct anaphora and indirect anaphora resolution. For direct anaphora resolution (same head noun) the algorithm achieves a recall of 79.9% and a precision of 84.4%; for indirect anaphora resolution a recall of 60% and a precision of 63.3% are achieved. The overall results of DD anaphora resolution achieve a recall of 75.3% and a precision of 79.3%. In general, the treatment of DD (identification and resolution) achieves a recall of 83.9% and a precision of 88.8%.

After the study of corpus, we observed that most of the time DDs headed by a demonstrative refer to an antecedent (Demonstrative DD's usually are anaphoric). It is evident that a demonstrative is used to refer to something previously introduced. Only in a dialogue can it be used to introduce a new discourse entity (deixis phenomena).

3.3 Comparison with other approaches

Several projects conducted by different authors show that most DD's add new discourse entities (have not referential properties). We emphasize works developed by Vieira & Poesio (Vieira & Poesio 98) and Bean & Riloff (Bean & Riloff 99) for English. The former demonstrates that about the 50% of DD's introduce a new entity. The latter shows that 63% of them introduce a new entity. These works use different corpus (33 articles from WSJ and the Latin American newswire articles from MUC-4, respectively). Moreover, the work developed by Muñoz et al. (Muñoz *et al.* 00) for Spanish using Lexesp corpus shows that 52% of DD's do not refer to any previous noun phrase. Obviously, the amount of DDs with no referential properties depends on the corpus and the language but the different work concludes that more than a half of DD introduce a new discourse entity. Consequently, algorithms without prior classifica-

C[a]	Non-anaphoric DD				Direct Anaphora				Indirect Anaphora			
	S[b]	E[c]	R[d]%	P[e]%	S	E	R%	P%	S	E	R%	P%
L[f]	429	20	89.9	95.5	117	20	79	85.4	32	16	65.3	66.6
D[g]	182	18	86.6	91	95	20	79.2	82.6	13	7	59.1	65
N[h]	99	11	84.6	90	42	7	84	85.7	12	10	50	54.5
	710	49	88.3	93.5	254	47	79.9	84.4	57	33	60	63.3

Overall			
S	**E**	**R%**	**P%**
311	80	75.3	79.5

Table 1: Anaphora resolution results after applying to test corpora

[a]C=Corpus
[b]S=Success
[c]E=Errors
[d]R=Recall
[e]P=Precision
[f]L=LEXESP
[g]D=DEED
[h]N=NEWSPAPER

tion (anaphoric or non-anaphoric) of DDs trying to solve the reference of all DD spend a lot of unnecessary computational time. Moreover, algorithms with prior classification of DDs that fail to accomplish in this task will be carried out even to the reference resolution task.

We classify coreference algorithms in two categories: algorithms that make no attempt to identify non-anaphoric DD prior to coreference resolution, and those that identify anaphoric and non-anaphoric DD before coreference resolution. We emphasize two algorithms belonging to the former group: Viera and Poesio's algorithm (Vieira & Poesio 98; Vieira 99) is based on several tests in order to classify the DDs as non-anaphoric expressions or to provide the correct antecedent. This algorithm achieved a recall of 62% and a precision of 83% in solving direct anaphora. Bridging descriptions were evaluated by hand, and 61 relations of 204 in the corpus were achieved. Moreover, the system achieved a recall of 74% and a precision of 85% for identifying new discourse entities (non-anaphoric DD). Bean and Rillof (Bean & Riloff 99) proposed a corpus-based mechanism to identify non-anaphoric DD. This algorithm uses statistical methods to generate lists of non-anaphoric DD (called *existencial*) and DD patterns from a training corpus. This algorithm for identifying non-anaphoric NPs achieves a recall of 78% and a precision of 86%.

Others algorithms without previous classification of DDs are: McCarty and Lehnert (McCarthy & Lehnert 95), who developed the system called RESOLVE using the C4.5 decision tree system (Quinlan 92). This system achieves a precision of 87.6% and a recall of 85.4%. Kameyama's algorithm (Kameyama 97), which described a pronoun and DD processing algorithm. This algorithm is applied to the information extraction system FASTUS (Hobbs *et al.* 96), and achieved a 46% recall for DD, a 62% recall for pronouns and a 69% recall for proper nouns.

Table 2 shows the results achieved by Vieira & Poesio and our algorithm for each task. Our algorithm achieves better results than Vieira & Poesio algorithms, but it can not be directly compare because different languages, different corpus and a different type of coreference are treated.

4 Conclusion

We have developed a semantic-driven algorithm for Spanish definite description resolution. The algorithm classifies DD into anaphoric or non-anaphoric, and it resolves direct anaphora and indirect anaphora. On

Task	V&P		Our Alg.	
	R%	P%	R%	P%
New Entity	74	85	88.3	93.5
Direct Anaphora	62	83	79.9	84.4
Bridging reference	52	46	60	63.3

Table 2: Indirect comparative of different task

the one hand, the algorithm is based on the automatic generation of a semantic network from the base concept of head nouns. These head nouns are extracted from all NPs that appear in the text and their base concepts are extracted using the Spanish Wordnet ontology. The mechanism only uses semantic information to make a previous classification of DD (anaphoric or not). The results obtained for the above mechanism show no high scores: only a 40% of non-anaphoric description can be identified using exclusively ontological concepts. However, if we use other semantic information such as synonym, hypernym, hyponym relation (restriction R1 & R2) the correct identification of non-anaphoric DD increases the precision to 93.5%. On the other hand a set of semantic preferences is applied to candidates in order to provide the correct antecedent, achieving a precision of 79.3% and a recall of 75.3%.

References

(Bean & Riloff 99) D. L. Bean and E. Riloff. Corpus-based Identification of Non-Anaphoric Noun Phrases. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistic*, pages 373–380, College Park, Maryland, USA, 1999.

(Clark 77) H. H. Clark. Bridging. In P. Johnson-Laird and P Wason, editors, *Thinking: readings in cognitive science*, pages 411–420. Cambridge University Press, London New York, 1977.

(Hobbs *et al.* 96) J. R. Hobbs, D. E. Appelt, J. Bear, D. Israel, M. Kameyama, M. Stickel, and M. Tyson. FASTUS: A cascaded finite-state transducer for extracting information from natural-language text. In E. Roche and Y. Schabes, editors, *Finite State Devices for natural Language Processing.* MIT Press, Cambridge, Massachussetts, 1996.

(Kameyama 97) M. Kameyama. Recognizing Referential Links: An Information Extraction Perspective. In Mitkov, R. and Boguraev, B., editor, *Proceedings of Association for Computational Linguistics and European Chapter Association for Computational Linguistics (ACL/EACL) Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, pages 46–53, Madrid, Spain, 1997.

(Llopis *et al.* 98) F. Llopis, R. Muñoz, A. Suárez, and A. Montoyo. EXIT: Propuesta de un sistema de extracción de información de textos notariales. *Novática*, 133:26–30, 1998.

(McCarthy & Lehnert 95) J. F. McCarthy and W. G. Lehnert. Using Decision Trees for Coreference Resolution. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence*, pages 1050–1055, 1995.

(Mitkov 97) R. Mitkov. Factors in anaphora resolution: they are not the only things the matter. A case study based on two different approaches. In ACL, editor, *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL/EACL)*, pages 14–21, Madrid, Spain, 1997.

(Montoyo & Palomar 01) A. Montoyo and M. Palomar. Specification Marks for Word Sense Disambiguation: New Development. In A. Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, volume 2004 of *Lecture Notes in Computer Science*, pages 182–191, Mexico City, 2001. Springer-Verlag.

(Muñoz *et al.* 00) R. Muñoz, M. Palomar, and A. Ferrández. Processing of Spanish Definite Descriptions. In O. Cairo, E. L. Sucar, and F. J. Cantu, editors, *Proceeding of Mexican International Conference on Artificial Intelligence*, volume 1793 of *Lectures Notes in Artificial Intelligence*, pages 526–537, Acapulco, Mexico, 2000. Springer-Verlag.

(Quinlan 92) J. R. Quinlan. *C4.5: Programs for Machine Learning.* Morgan Kaufmann, San Mateo, CA, 1992.

(Vieira & Poesio 98) R. Vieira and M. Poesio. Processing definite descriptions in corpora. In S. Botley and T. McEnery, editors, *Corpus-based and computational aproach to anaphora.* UCL Press, London, 1998.

(Vieira 99) R. Vieira. Co-reference resolution of definite descriptions. In *Proceedings of VI Simposio Internacional de comunicación Social*, pages 497–503, Santiago de Cuba, Cuba, 1999.