Solving Definite Descriptions through Dialogue Structure in Spanish

Borja Navarro, Patricio Martínez-Barco and Rafael Muñoz
Departamento de Lenguajes y Sistemas Informáticos.
Universidad de Alicante.
Alicante, Spain
{borja,patricio,rafael}@dlsi.ua.es

Abstract

In this paper we present an study about the relationship between the definite description resolution and the structure of dialogues defining an anaphoric accessibility space. This relationship allows to reduce the list of candidates in the resolution process. This anaphoric accessibility space is built with a series of open sequences where the coreference is likely used. The use of this anaphoric accessibility space reduce both the computational time and the possibility of obtaining an incorrect antecedent in the resolution process. Moreover, the definition of this anaphoric accessibility space based on dialogue structure only depends on the self structure.

1 Introduction

Dialogue systems constitute a very exploited group of applications in natural language processing. Nevertheless, until a few years ago, this kind of systems were developed as isolated domain dependant systems. Nowadays, there is an increasing interest in obtaining NLP resources providing the basis for generic dialogue systems that can be applied to whatever domain with only performing minor changes in some of their modules.

According to this, in Allen *et al.* (2000) a generic architecture for dialogue systems is described. This architecture is based on the use of several modules, mostly of them domain independent, and some of them domain dependent but easily adaptable to whatever domain. According to Allen *et al.* (2000) , one of the most critical domain independent modules in dialogue systems is the Dialogue Manager (also called the Interpretation Manager). This module is responsible for the interpretation task: it coordinates a range of processes to recognize the user's intentions underlying the utterance and to compute new discourse obligations.

In this way, one of the domain independent modules that is invoked by the Discourse Manager is the Reference Manager. The Reference Manager must be domain independent in order to be easily adapted to whatever dialogue system and it will attempt to identify likely referents for referring expressions (coreference resolution). The Reference Manager must use the accumulated discourse context from previous utterances plus knowledge of the particular situation to identify candidates.

Previous work about coreference resolution showed several linguistic and statistical rules that had been adopted in order to define the suitable candidate in each situation. These rules involved morphologic, syntactic and semantic information. However, our state is that also information about dialogue structure must been used in order to solve the coreference in dialogues.

In this paper we will present a nominal anaphora resolution system that solves the coreference due to definite noun phrases (definite descriptions) in dialogue systems. We will focus on direct anaphora resolution (where the anaphoric expression has the same head than its antecedent or the head is omitted).

Our system is based on the algorithm presented by Muñoz and Palomar (2001) to solve references in definite descriptions applied to monologues. The algorithm uses linguistic information - morphological, syntactical and semantics - to establish the core-

ferentiality conditions between the referent and the referring expression. In order to apply this algorithm to dialogues it has been extended by means of the incorporation of an anaphoric accessibility space according to previous work about anaphora resolution in dialogues performed in (Palomar and Martínez-Barco, 2001).

The organization of the paper is as follows: section 2 presents the main features of the definite description algorithm, section 3 shows the development about anaphoric accessibility space in dialogues performed by the authors, section 4 shows the integration the anaphoric accessibility space in our algorithm, and finally, section 5 and 6 shows the results of the experimental work developed and the conclusions, respectively.

2 Definite description algorithm

The algorithm contains the following main components.

1. Clustering module. In this module an algorithm to cluster DDs is applied to classify them into anaphoric or non-anaphoric. The clustering task uses the EuroWordNet's ontology instead of a distance.

2. Coreference module. This module uses a set of constrains and preferences to provide the correct antecedent. The following sets of rules are applied:

   - A set of semantic constraints that rule out anaphoric DD-NP dependence.
   - A set of preferences obtained from a empirical study is applied using a weight management system. 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 Clustering task

The algorithm goes through the text looking for noun phrases. Once, a noun phrase is detected, his head noun is extracted. The head noun is used to obtain from EuroWordNet's ontology his base concept. The clustering technique uses this base concept instead of a distance to classify the noun phrases into equivalent classes. If there are at least one noun phrase belonging to the same class, then noun phrases are semantically compatible. Moreover, if the noun phrase founded in the text is a DD then it is classified provisionally as anaphoric and the resolution algorithm is applied. Otherwise, if there are no noun phrase belonging to the same semantic class then the DD is classified as non-anaphoric and the algorithm goes on.

Furthermore, taking advantage of the clustered noun phrases into semantically equivalent classes, the amount of comparison between antecedents and the DD has been reduced although the solution searching space used by the resolution algorithm that is made up by all previous sentences. This fact is due to that a DD is only compared with noun phrases belonging to the same class, because head nouns semantically related through synonym, hypernym or hyponymy relations are semantically equivalent.

The mechanism is based on a simple idea: a DD will be non-anaphoric if it is not semantically compatible. 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.

A Word Sense Desambiguation module is needed in order to provide the correct sense of head nouns. The Specification Marks module (Montoyo and Palomar, 2001) 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 cluster 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 semantically compatible NPs is made.

If there is an antonym relationship (left - right) between two modifiers then the NP is rejected.

2.3 Preference Management

The system scores a salience value for each possible antecedent no rejected by constrains. The antecedent with highest salience value is chosen as antecedent. For each candidate, preferences are applied adding the weight of the fulfilled ones to the salience value of the candidate.

A set of preferences obtained from a empirical study is applied. For each preference, a set of values are assigned to several salience parameters (frequency of mention, proximity, semantic relations) for an NP.

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

P2 Pre and post-modifiers relation. The system selects antecedents 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 antecedents 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 the NP with the same head noun and later with a semantic relationship between head nouns.

P5 Gender and number agreement. The system selects the antecedent with gender and number agreement.

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

[1]Definite descriptions with the different head noun as their antecedent were called bridging references by Clark (Clark, 1977)

The algorithm does not limit the number of previous sentences to be use in order to search the correct antecedent of 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 reducing the number of comparison. Moreover, as it will be shown below, this searching space could be reduced through the definition of an anaphoric accessibility space.

3 An annotation scheme for dialogue structure

For successful anaphora resolution in dialogues, we assume that it is essential to identify dialogue structure (Palomar and Martínez-Barco, 2001). Therefore, we propose an annotation scheme for Spanish dialogues that is based on work carried out by Gallardo (1996), who applies the theories put forward by Sacks *et al.* (1974) concerning conversational turn-taking.

We use an annotation scheme based on these theories for three main reasons. First, as it is a general approach to dialogue modeling, it is applicable to all types of dialogues, including both task-oriented and information-retrieval-oriented dialogues. Consequently, the use of such a model as a basis for developing our anaphor resolution procedure allows us to apply the procedure to any type of domain, thus offering an advantage over procedures based on discourse models specific to particular domains. Second, this annotation scheme can be easily applied to automatic processes without metalinguistic considerations. Although in our work the annotation task has been performed by hand, for dialogue-based applications in which our procedure might be embedded (e.g., in dialogue management systems), annotation tasks must be performed automatically. Finally, we wanted to base our own procedure on studies of the influence of dialogue structure on anaphora resolution that were carried out by Fox (1987), whose approach, in turn, is based on that of Sacks *et al*.

According to these theories, the basic unit of conversation is the *move*, which informs the listener about an action, request, question, etc. Moves are carried out by means of *utterances*.[2] And utterances

[2]An *utterance* in a dialogue is equivalent to a sentence in a non-dialogue, although, because of the lack of punctuation

are joined together to become *turns*.

Since our work was done using spoken dialogues that had been transcribed, turns are annotated in the texts and utterances are delimited by the use of punctuation marks or by the ends of turns. Reading a punctuation mark (., ?, !, ...) allows us to recognize the end of an utterance. These tasks do not affect the anaphora-resolution process.

As a result, we propose the following annotation scheme for dialogue structure:

Turn (T) is identified by a change of speaker in the dialogue; each change of speaker presupposes a new turn. On this point, we make a distinction between two different kinds of turns:

- An **intervention turn (IT)** is one that adds information to the dialogue. Such turns constitute what is called *the primary system of conversation*. Speakers use their interventions to provide information that facilitates the progress of the topic of conversation. Interventions may be **initiatives (IT$_I$)** when they formulate invitations, requirements, offers, reports, etc., or **reactions (IT$_R$)** when they answer or evaluate the previous speaker´s intervention. Finally, they can also be **mixed interventions (IT$_{R/I}$)**, which is a reaction that begins as a response to the previous speaker's intervention, and ends as an introduction of new information.
- A **continuing turn (CT)** represents an empty turn, which is quite typical of a listener whose aim is the formal reinforcement and ratification of the cast of conversational roles. Such interventions lack information.

Adjacency pair (AP) (also called **exchange**) is a sequence of turns headed by an initiation intervention turn (IT$_I$) and ended by a reaction intervention turn (IT$_R$). This form of anaphora, in which the reference appears within an adjacency pair, appears to be very common in dialogues (Fox, 1987).

Topic (TOPIC). The topic must be a lexical item that is referred to frequently. According to Rocha (1998), four features are taken into account in the selection of the best candidate for a discourse topic: frequency, even distribution, position of first token, and semantic adequacy. A highly frequent element that occurs intensively in a passage of the dialogue but does not appear for long stretches is not likely to be a good choice for discourse topic. In the same way, neither is an element whose first appearance occurs a long way from the beginning the best choice. Moreover, semantic adequacy must be considered for the candidate, and it must be assessed by the annotator.

Based on the above-mentioned structure, then, the following tags are considered necessary for dialogue structure annotation: **IT$_I$**, **IT$_R$**, **CT**, **AP**, and **TOPIC**. The AP and TOPIC tags will be used to define the anaphoric accessibility space, and the remaining tags will be used to obtain the adjacency pairs. The IT$_{R/I}$ tag, representing mixed interventions, is not included since mixed interventions can be annotated as IT$_R$ plus IT$_I$. This task is done in the annotation phase.

4 Definite description resolution in dialogues

Based upon the above-mentioned annotation, in Palomar and Martínez-Barco (2001), an anaphoric accessibility space was proposed for Spanish in order to resolve anaphors in the form of personal and demonstrative pronouns.

That proposal was based on previous work by Fox (1987), who asserted that the first mention of a referent in a sequence of contexts is performed with a full noun phrase. After that, by using an anaphor the speaker displays an understanding that sequence has not been closed down.

To build an anaphoric accessibility space, Palomar and Martínez-Barco performed an study of the different sequences that could be open when an anaphor appears. These sequences were the following:

- the adjacency pair containing the anaphor, plus

marks, utterances are recognized by means of speakers' pauses.

- the adjacency pair preceding the adjacency pair containing the anaphor, plus
- any adjacency pair including the adjacency pair containing the anaphor, plus
- the noun phrase representing the main topic of the dialogue.

The anaphoric accessibility space proposed in Palomar and Martínez-Barco (2001) showed successful results when it was applied together with a pronominal anaphora resolution algorithm. According to their proposal, the algorithm looked for the solution in that space, discarding solutions out of those sequences. Furthermore, an adequate ordering of those sequences was used to improve the preference system used giving different importance to solutions appearing in each kind of sequence. Authors showed an improvement of 20% when the anaphoric accessibility space was incorporated.

Our aim in this paper is to show how the incorporation of an anaphoric accessibility space based on the dialogue structure helps to the definite description resolution system just as it helped to pronoun resolution. That is to detect those cases in which this space could reduced the solution searching space, and then incorporate it into the algorithm.

For this reason, we have performed the experimental work showed in the next section.

5 Experimental work

Data for this experimental work were taken from the *Corpus InfoTren: Person*, a corpus of 204 transcribed spoken Spanish dialogues provided by the Basurde Project (Proyecto BASURDE, 1998 2001). These dialogues are conversations between a railway company employee and a client. The transcriptor used in the Basurde Project provides turn and speaker markup.

5.1 Definite Description analysis

In our corpus, we have focused our interest in the next kind of definite descriptions[3]:

1. Definite descriptions that have a relation of repetition with their antecedent: This is the most common kind of definite description in the corpus (73%).

   For example:

   OP: tiene a las seis en punto un Euromed, luego a las siete de la tarde un Estrella (...), a las siete y media **un Talgo** (...)

   you have an Euromed at six o'clock, then an Estrella at seven o'clock (...), ***a Talgo*** *at half past seven (...)*

   US: sí (...)

   yes (...)

   OP: **el Talgo** de las diecinueve treinta (....)

   the Talgo *at nineteen thirty (...)*

   We have found 283 definite descriptions with a relation of repetition with their antecedent. We have split this class of definite description in different sub-classes:

   (a) Definite descriptions that are identical to their antecedent (head noun and determiner): 253 definite descriptions.

   For example,

   US: o sea, **el Euromed** son tres horas

   that is, ***the Euromed*** *spends three hours*

   OP: **el Euromed** tres horas (...)

   the Euromed *spends three hours (...)*

   OP: ¿este?, a ver... **el Euromed** son cuatro mil seiscientas"

   this one?, ... ***The Euromed*** *costs four thousand and six hundred*

   (b) Definite descriptions that repeat the head noun of the nominal phrase antecedent: 25 definite descriptions.

   For example,

   OP: hay **una cabina** (...)

   there is ***a cabin*** *(...)*

   US: ¿qué vale **la cabina** de dos camas?"

   how much is ***the cabin*** *with two beds*

   (c) Definite descriptions that only repeat the lemma of the head noun of the antecedent (normally, in this sub-class, changes in the

[3] About the different classifications of definite descriptions, see (Vieira and Poesio, 2000), (Muñoz et al., 2000),(Navarro et al., 2001)

part of speech and/or changes in the derivational morphology of the word is produced): 5 definite descriptions.

For example,

US: sí, vale, ¿y para **reservar**?
*yes, ok, and **to book**?*

OP: sí, pues **las reservas** en este mismo número las hacemos (...)"
*yes, then you can make **the booking** in this same number (...)*

There is a particular class of definite descriptions that repeat completely their antecedent: those ones produced to ensure the correctly interpretation of the words. These repetitions no contribute to the development of the conversation, and these repetitions are frequent in speech dialogues. In a strict sense, these are not anaphoric definite descriptions.

For example,

US: y **el coche** ¿cuánto vale?
*and **the car**, how much is it?*

OP: entonces, **el coche** ¿qué coche lleva?
*then, **the car**, what car do you drive?*

US: ¡huy madre! Un Audi (...)
Oh my good! an Audi (...)

OP: bueno, si es mediano
well, if it is medium-size

US: si mediano es
yes, it is medium-size

OP: si es mediano, le vale a Madrid diez mil (...)
yes, if it is medium-size, the cost to Madrid is ten thousands

US: el coche
the car

OP: el coche
the car

2. Definite descriptions that have a lexical relation (synonymy, hyponymy, hypernymy, and so on) with his antecedent.

   For example:

   US: serían dos adultos y **un niño** (...)
   *it would be two adults and **a child** (...)*

   OP: si quieren ir los dos solos con **el bebé** en una cabina (...)
   *if you want to go alone both with **the baby** in a cabin (...)*

   We have found only 15 of this kind of definite descriptions. This is due to the particular features of the corpus producing that speakers do not use this kind of definite descriptions frequently.

3. Definite descriptions that have a elliptic head noun (or adjectival anaphor). In this kind of definite description, the head of the nominal phrase is elliptic. The phrase consists only in a determiner and an adjectival or prepositional phrase.

   For example:

   OP: ¿qué quiere ir, **en cabina de cuatro**, de dos o de uno?
   *what do you want to go, in a **four**, two or one people **cabin**?*

   US: depende del precio, a ver
   it depends on the price, let me see

   OP: la de cuatro vale nueve mil pesetas (...)"
   ***the four one** costs nine thousand pesetas (...)*

We have found 90 definite descriptions having elliptic head noun.

5.2 Accessibility space components

We have established five different structural anaphoric accessibility space components:

1. Same adjacency pair: the anaphoric expression and its antecedent are located in the same adjacency pair.

2. Previous adjacency pair: the antecedent is located in the previous adjacency pair.

3. Nested adjacency pair: the antecedent is located in a higher level adjacency pair, that includes the adjacency pair of the definite description.

4. Topic of discourse: the antecedent is, directly, the topic of discourse.

5. Beyond the previous adjacency pair: finally, the antecedent of a definite description can be located beyond the previous adjacency pair.

5.3 Relations between definite description and accessibility space

With this, we can investigate the relation and influence of the anaphoric definite description and the structure of dialogue. For this, we will specify the structural anaphoric accessibility space for each kind of definite description.

1. Repetition relation. The definite descriptions that have a complete repetition of their antecedents locate them in the same adjacency pair in 82 cases (28.77%), in the previous adjacency pair in 48 cases (16.84%) and in a nested adjacency pair in 20 cases (7.01%). Finally, in 75 cases (26.31%), the antecedent is located beyond the previous adjacency pair.

   Regarding definite descriptions that have a partial repetition with his antecedent (repetition of head noun), the situation is similar. In 16 cases, the antecedent is located in the same adjacency pair, and in 9 cases the antecedent is located beyond the previous adjacency pair.

   Finally, we have found few definite descriptions that repeat the lemma of the antecedent (6 cases). In this cases, the antecedent is located in the same adjacency pair.

2. Lexical relation. The definite descriptions that have a lexical relation with his antecedent locate it in the previous adjacency pair en 4 cases (1.40%) and they locate his antecedent beyond the previous adjacency pair in 10 cases (3.50%). In our corpus, no definite description with a lexical relation with their antecedent have been located it in the same adjacency pair. The reason is that there are very few definite descriptions with this kind of relation in corpus.

3. Elliptical definite descriptions. Regarding definite descriptions that have a elliptical head noun, their antecedents are located in the same adjacency pair in 19 cases (6.6%), in the previous adjacency pair in 35 cases (12.20%), in a nested adjacency pair in 4 cases (1.28%), in topic in 13 cases (4.56%), and, finally, the antecedent is located beyond the previous adjacency pair in 23 cases (8.07%).

5.4 Discussion

In general, this result shows that the antecedent of the definite descriptions are located in different zones of the text structure (Ariel 1990) and, in our investigation, in different zones of the dialogue structure. For this, the information about dialogue structure is important in correference resolution.

For definite descriptions with a repetition relation with their antecedent, the same adjacency pair is the main zone of dialogue structure where the antecedent is located. Normally, this repetition is produced in the answer turn of a question turn. For example:

US: ¿me dice **el precio**?"
*Can you tell me **the price**?*
OP: **el precio** son tres mil ochocientas (...)"
***the price** is three thousand and eight hundred (...)*

In this kind of definite descriptions, in many cases the antecedent is located in previous adjacency pair or in nested adjacency pair. Normally, this repetition is used to confirm the information in the dialogue. For example:

OP: Hasta **el último** que es el de las veintidós ocho
*... until **the last one** that is the one of the twenty two eight*
US: veintidos ocho, **el último**
*twenty two eight, **the last one***
OP: sí (...)"
yes (...)
US: y, ¿**la vuelta**?
*and, **the return**?*
OP: ¿**la vuelta** en Euromed desde Valencia?
***the return** on Euromed from Valencia?*
US: sí
yes
OP: a las siete cero cinco (...)"

at seven zero five (...)

Finally, the definite description with a relation of repetition with its antecedent can locate it beyond the previous adjacency pair. Unlike pronouns, the definite descriptions are not empty of sense, so they have more semantic information in order to find his antecedent. So a definite description can locate its antecedent beyond the previous adjacency pair, because they have sufficient information to find his antecedent.

As regards definite descriptions that have a lexical relation with their antecedents, we have not found an enough number of cases in order to achieve a conclusion. However, they are similar to previous definite descriptions.

Finally, there are a lot of definite descriptions with head noun elliptic (or adjectival anaphor). In this dialogues, the topic is often the same: information about trains and timetables. With that, it is normal that the interlocutors refer to topic with a definite description that has the head noun elliptic. The referent is often the topic of the dialogue or other definite description that refers to topic. For example:

OP: Hay un tren a las siete y media (...)
There is a train at half past seven (...)

OP: **El de las siete media** llega a las nueve (...)"
the one at half past seven *arrives at nine o'clock (...)*

In general, the definite descriptions that have a elliptical head noun are used to organize the general structure of the dialogue. In a first turn, the speaker introduce a number of entities (normally, the trains), and, afterward, the interlocutor refers to each one with definite description with elliptical head noun.

6 Conclusion

In this paper an study that shows the relationship between the definite description resolution and the structure of dialogues has been presented. This relationship allows to reduce the list of candidates in the resolution process. For this reason, we have presented an algorithm in which we have integrated an anaphoric accessibility space from which the candidates are going to be extracted. This anaphoric accessibility space is built with a series of open sequences where the coreference is likely used.

The use of this anaphoric accessibility space reduce both the computational time and the possibility of obtaining an incorrect antecedent in the resolution process. Moreover, the definition of this anaphoric accessibility space based on dialogue structure does not depend on a prefixed number of sentences such as proposed by other authors (that is obviously corpus-dependent), but it only depends on its own structure.

However, there are some characteristics in the definite descriptions of the corpus that must be discussed. First, there are some undefined nominal phrases that refers directly to topic. Second, there are some definite descriptions that have multiple antecedents: normally, a speaker introduces some referents, and afterwards he refers to them as a whole. Both kinds of noun phrases has not been treated in this paper. Besides, definite descriptions having pragmatic relationship with their antecedent are not treated due to the lack of resources providing this information.

7 Acknowledgement

This paper has been supported by the Spanish Government (MCYT) under grant TIC2000-0664-C02-01/02.

References

J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent. 2000. An Architecture for a Generic Dialogue Shell. *Natural Language Engineering*.

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

B. Fox. 1987. *Discourse Structure and Anaphora.* Written and conversational English. Cambridge Studies in Linguistics. Cambridge University Press, Cambridge.

B. Gallardo. 1996. *Análisis conversacional y pragmática del receptor*. Colección Sinapsis. Ediciones Episteme, S.L., Valencia.

A. Montoyo and M. Palomar. 2001. 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. Springer-Verlag.

R. Muñoz and M. Palomar. 2001. Clustering technique based on semantic for definite description resolution. In *Proceedings of the International Conference Text Speech and Dialogue*, volume 2166 of *Lecture Notes in Artificial Intelligence*, Czech Republic. Springer-Verlag.

R. Muñoz, M. Palomar, and A. Ferrández. 2000. 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. Springer-Verlag.

B. Navarro, P. Martínez, and R. Muñoz. 2001. Propuesta de un espacio de accesibilidad anafórica estructural para textos HTML. *Procesamiento del Lenguaje Natural*, 27:97–106.

M. Palomar and P. Martínez-Barco. 2001. Computational approach to anaphora resolution in Spanish dialogues. *Journal of Research in Artificial Intelligence*, 14.

Proyecto BASURDE. 1998-2001. Spontaneus-Speech Dialogue System in Limited Domains. Comisión Interministerial de Ciencia y Tecnología TIC98-423-C06. http://gps-tsc.upc.es/ veu/basurde/Home.htm.

M. Rocha. 1998. *A corpus-based study of anaphora in dialogues in English and Portuguese*. Ph.D. thesis, University of Sussex, Sussex. UK.

H. Sacks, E. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn taking for conversation. *Language*, 50(4):696–735.

R. Vieira and M. Poesio. 2000. An Empiricall Based System for Processing Definite Descriptions. *Computational Linguistics*, 26(4):539–593.