# An annotation scheme for dialogues applied to anaphora resolution algorithms

Patricio Martínez-Barco and Manuel Palomar

Departmento de Lenguajes y Sistemas Informáticos
Universidad de Alicante
Carretera de San Vicente del Raspeig - Alicante - Spain
Tel. +34965903653 Fax. +34965909326
{patricio, mpalomar}@dlsi.ua.es

**Abstract.** Usually, anaphora resolution researchers focus their works on defining a set of sources to extract useful information that helps to anaphora resolution. Other works base their research on extracting relationships between utterances and anaphors obtaining satisfactory results working on languages with short anaphoric accessibility spaces like English. In this work we state that anaphora resolution in Spanish needs an adequate definition of accessibility spaces and then we propose an annotation scheme for dialogue structure to define these spaces. This proposal has been tested achieving successful results when it is applied to real man-to-man telephone conversations.

## 1 Introduction

In this paper, an annotation scheme for dialogues applied to anaphora resolution algorithms is presented. This annotation scheme for Spanish dialogues is based on the theories about the taking of speaking turns. This scheme is oriented to the definition of an anaphoric accessibility space where the antecedents of an anaphor can be found. The adequate definition of this space allows to optimize the anaphora resolution algorithm.

According to Dahlbäck [4], the efforts made so far towards resolving anaphora can be divided into two basic approaches : Traditional and Discourse-oriented. The traditional approach generally depends on linguistic knowledge. In the discourse-oriented approach, however, the researcher tries to model the complex structure of discourse. Anaphora, accepted as discourse phenomena, is resolved with the help of that complex structure. These works are mostly focused on defining anaphora resolution algorithms, both the traditional approaches (Hobbs [8], Baldwin [1], Mitkov [11]) and the discourse-oriented ones (Grosz et al [7], Strube [17]).

However, the former do not perform a defined proposal about anaphoric accessibility space, and the latter constraint the space for possible antecedents to the previous utterance. Although, this strategy is adequate for English processing, its application to other languages such as Spanish is not such suitable.

For instance, Spanish personal pronouns contain more morphological information than English ones. This makes Spanish speakers to expect larger anaphoric accessibility spaces. Also, Spanish free-word-order causes problems when grammatical role information must be used.

Following the idea that any treatment of anaphora must consider the structure of its source text-types, Fox [5] investigated anaphora looking for an adequate accessibility space on the basis of the relationship between discourse structure and this phenomena. In this work, Fox used the theories of conversational analysis by Sacks, Schegloff and Jefferson [15] to define the structure of discourse. According to this approach, the adjacency pair is the basic structural unit of dialogues. Thus, a close relationship between adjacency pairs and anaphoric accessibility space is shown.

In the next section an annotation scheme for Spanish dialogue structure is proposed. After that a brief survey of the anaphora resolution algorithm is shown, and finally, the evaluation of the annotation scheme applied to the anaphora resolution algorithm is presented.

## 2 An annotation scheme for dialogue structure

For the successful processing and resolution of anaphora in dialogues, we believe that the proper annotation of the dialogue structure is necessary. With such a view, we propose an annotation scheme, for Spanish dialogues, that is based on the work carried out by Gallardo [6], who applies, to Spanish dialogues, the theories put forward by Sacks, Schegloff and Jefferson [15] about the taking of speaking turns (conversational). According to these theories, the basic unit of knowledge is the *move* that can inform the listener about an action, request, question, etc. These moves are carried out by means of *utterances*. Therefore, utterances are joined together to become *turns*.

Since our work was done on spoken dialogues that have been written (transcribed), the turn appears annotated in the texts and the utterances are delimited by the use of punctuation marks. The reading of a punctuation mark (., ?, !, ...) allows us to recognize the end of an utterance.

As a conclusion, therefore, we propose the following annotation scheme for dialogue structure based on the structure defined by Gallardo [6]. This proposal is justified on the studies by Fox [5] about discourse structure and anaphora resolution:

**Turn (T)** is identified by a change of speaker in the dialogue; each change of speaker supposes a new speaking turn. On this point, Gallardo makes 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** ($\mathbf{IT}_I$) when they formulate invitations, requirements, offers, reports, etc., or **reactions** ($\mathbf{IT}_R$) when they answer or evaluate

the previous speaker's intervention. Finally, they can also be **mixed interventions ($\mathbf{IT}_{R/I}$)**, meaning 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** or **Exchange (AP)** is a sequence of turns T headed by an initiation intervention turn ($\mathrm{IT}_I$) and ended by a reaction intervention turn ($\mathrm{IT}_R$). One form of anaphora which appears to be very common in dialogues is the reference within an adjacency pair [5].

**Topic (TOPIC)** is the main entity in the dialogue. According to Rocha [14] four features are taken into account in the selection of the best candidate for discourse topic: frequency, even distribution, position of first token, and semantic adequacy. The topic must be a lexical item which is frequently referred to.

According to the above-mentioned structure, the following set of tags is considered necessary for dialogue structure annotation: $\mathrm{IT}_I$, $\mathrm{IT}_R$, $\mathrm{IT}_{R/I}$, CT, AP and TOPIC. AP and TOPIC tags will be used to define the anaphoric accessibility space and the remaining will be used to obtain the adjacency pairs.

For this experiment, the corpus has been manually annotated. However, nowadays there are some works performing an automatic adjacency pair tagging, such as the *Basurde* project [2]. On the other hand, there are other works performing automatic topic tagging (e.g. Reynar [13]) or automatic topic extraction (e.g. the method for anaphora resolution shown in Martínez-Barco *et al* [9]).

The annotation of conversational dialogues is carried out, as shown above, and the evaluation of the proposed anaphoric accessibility space accomplished. An important aspect of this type of annotation is the training phase, which assures the reliability of the annotation.

The annotation phase is accomplished in the following way: a) two annotators are selected, b) an agreement is reached between the two annotators with regard to the annotation scheme using 5 dialogues (training corpus), c) the annotation is then carried out by both annotators in parallel over the remaining 35 dialogues (test corpus) and, d) finally, a reliability test is done on the annotation (see Carletta *et al.* [3]) in order to guarantee the results. The reliability test uses the *kappa* statistic that measures the affinity between the annotations of the two annotators by making judgements about categories. See Siegel and Castellan [16] for *kappa* statistic ($k$) computing.

According to Carletta, a $k$ measurement such as $0.68 < k < 0.8$ allows us to make encouraging conclusions, and $k > 0.8$ means total reliability between the results of both annotators.

Once both annotators have carried out the annotation, the reliability test of the annotation has been run, with a *kappa* measurement of $k = 0.91$. We therefore consider the annotation obtained for the evaluation to be totally reliable.

## 3 A brief survey of the algorithm

Based on the above-mentioned dialogue structure, an algorithm that combines several kinds of knowledge must be proposed. Concretely, it must use and combine: a) linguistic knowledge such as lexical, morphological, syntactic and semantic knowledge; and b) knowledge from the dialogue structure itself, which is based on the annotation previously proposed, and knowledge taken from the topic of the dialogue.

Following the previous directives, an anaphora resolution algorithm for dialogues was proposed by Palomar and Martínez-Barco [12]. This algorithm was based on the following three steps:

1. All possible antecedents are obtained from the structure and topic of dialogue texts.
2. Linguistic constraints are applied on the previously selected antecedents, with the aim of discarding incompatible antecedents:
3. If there are more than one antecedent left after applying the previous constraints, then a set of linguistic and dialogue structure preferences are applied.

## 4 Evaluation of the annotation scheme

In order to carry out the evaluation of this scheme, we had 200 transcribed spoken dialogues (*Corpus InfoTren: Person*) at our disposal, afforded us by the project *Basurde*, from which we selected and POS-tagged 40. The tagging provides lexical and morphological knowledge for each word of the text. Afterwards, the structure of the dialogue was manually annotated, following the directives previously presented. From the 40 POS-tagged dialogues, 5 were randomly selected for the training of the algorithm and the remaining 35 were reserved in order to carry out the final evaluation. As a result, 73.8% precision for pronominal anaphora resolution and 78.9% for adjectival anaphora resolution was achieved.

An empirical study about the precision of the anaphoric accessibility space using this schema is shown in Martinez-Barco and Palomar [10].

## 5 Conclusions

In this paper we have proposed an annotation scheme of dialogue structure that helps to anaphora resolution in Spanish dialogues. This annotation scheme has been developed on the basis of works by Gallardo [6] who applied the theory of conversational analysis to Spanish dialogues and the studies by Fox [5] about the relationship between this theory and anaphora. Furthermore, an evaluation of this proposal over an anaphora resolution algorithm has been shown with successful results.

## 6 Acknowledgments

The authors wish to thank N. Prieto, F. Pla and A. Molina (Universitat Politecnica de Valencia) for having contributed their tagger and E. Segarra (Universitat Politecnica de Valencia) for affording us the *Corpus InfoTren: Person* from the BASURDE Project.

This paper has been supported by the Spanish Government under grant HB1998-0068.

## References

1. B. Baldwin. CogNIAC: high precision coreference with limited knowledge and linguistic resources. In *Proceedings of ACL/EACL workshop on Operational factors in practical, robust anaphora resolution*, Madrid (Spain), July 1997.
2. Proyecto BASURDE. *Spontaneus-Speech Dialogue System in Limited Domains*. CICYT (TIC98-423-C06), 1998-2001. http://gps-tsc.upc.es/veu/basurde/Home.htm.
3. J. Carletta, A. Isard, S. Isard, J.C. Kowtko, G. Doherty-Sneddon, and A.H. Anderson. The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics*, 23(1):13–32, 1997.
4. N. Dahlbäck. *Representations of Discourse-Cognitive and Computational Aspects*. PhD thesis, Department of Computer and Information Science, Linköping University, Linköping, Sweden, 1991.
5. B. Fox. *Discourse Structure and Anaphora*. Written and conversational English. Cambridge Studies in Linguistics. Cambridge University Press, Cambridge, 1987.
6. B. Gallardo. *Análisis conversacional y pragmática del receptor*. Colección Sinapsis. Ediciones Episteme, S.L., Valencia, 1996.
7. B. Grosz, A. Joshi, and S. Weinstein. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225, 1995.
8. J. Hobbs. Resolving pronoun references. In B. Grosz B.L. Webber and K. Jones, editors, *Readings in Natural Language Processing*. Morgan Kaufmann, Palo Alto, CA, 1986.
9. P. Martínez-Barco, R. Muñoz, S. Azzam, M. Palomar, and A. Ferrández. Evaluation of pronoun resolution algorithm for Spanish dialogues. In *Proceedings of the Venezia per il Trattamento Automatico delle Lingue (VEXTAL'99)*, pages 325–332, Venice (Italy), November 1999.
10. P. Martínez-Barco and M. Palomar. Empirical study of the anaphoric accessibility space in Spanish dialogues. In *Proceedings of the Fourth Workshop on the Semantics and Pragmatics of Dialogue (GÖTALOG'2000)*, Göteborg, Sweden, June 2000.
11. R. Mitkov. Robust pronoun resolution with limited knowledge. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL'98)*, Montreal (Canada), August 1998.
12. M. Palomar and P. Martínez-Barco. Anaphora Resolution through Dialogue Adjacency Pairs and Topics. In D.N. Christodoulakis, editor, *Natural Language Processing - NLP 2000*, volume 1835 of *Lecture Notes in Artificial Intelligence*, pages 196–203, Patras, Grece, June 2000. Springer-Verlag.

13. Jeffrey C. Reynar. Statistical Models for Topic Segmentation. In *Proceedings of 37th Annual Meeting of the Association for Computational Linguistics*, pages 357–364, Maryland, USA, June 1999.
14. M. Rocha. *A corpus-based study of anaphora in dialogues in English and Portuguese.* PhD thesis, University of Sussex, Sussex. UK, 1998.
15. H. Sacks, E. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn taking for conversation. *Language*, 50(4):696–735, 1974.
16. S. Siegel and J. Castellan. *Nonparametric Statistics for the Behavioral Sciences.* McGraw-Hill, 2nd edition, 1988.
17. M. Strube and U. Hahn. Functional Centering - Grounding Referential Coherence in Information Structure. *Computational Linguistics*, 25(5):309–344, 1999.