Procesamiento del Lenguaje Natural, num. 29 (2002), pp.223-230 recibido 24-05-2002; aceptado 03-07-2002

Discourse Marker characterisation via clustering: extrapolation from supervised to unsupervised corpora *

Laura Alonso
CLiC, Department of General Linguistics
Universitat de Barcelona
lalonso@lingua.fil.ub.es

Irene Castellón
Department of General Linguistics
Universitat de Barcelona
castel@lingua.fil.ub.es

Karina Gibert
Department of Statistics
and Operational Research
Universitat Politècnica de Catalunya
karina@eio.upc.es

Lluís Padró
TALP Research Center
Software Department
Universitat Politècnica de Catalunya
padro@lsi.upc.es

Resumen: En este artículo mostraremos como las técnicas de clustering pueden aportar evidencia empírica para una caracterización de los Marcadores del Discurso (DMs) que contribuya a superar la falta de consenso y reduzca el coste de construcción de los recursos de PLN basados en DMs. Hemos establecido una noción de prototipicalidad de DMs comparando las clasificaciones de corpus anotado manualmente y automáticamente, a partir de la cual podemos obtener clasificaciones fiables a partir de corpus anotado automáticamente.
Palabras clave: Marcadores del Discurso, Clustering, Discurso

Abstract: In this paper we will show how clustering techniques provide empirical evidence for a characterisation of Discourse Markers (DMs) that helps in overcoming the lack of consensus and reduces the cost of building NLP resources based on DMs. By comparison of classifications from hand-tagged and unsupervised corpora we are capable of grounding a notion of DM prototypicality, from which reliable classifications can be obtained from fully unsupervised corpora.
Keywords: Discourse Markers, Clustering, Discourse

1 *Motivation*

Some approaches to automated discourse processing rely on shallow textual clues to obtain a representation of discourse that is useful for a determined NLP task. Cue phrases such as *because*, *although* or *in that case*, usually called Discourse Markers (DMs), are among the most popular, because they are highly informative of discourse structure and they can be treated satisfactorily with shallow NLP techniques[1].

As a drawback to their low processing cost, work concerning DMs usually implies labour-intensive description and encoding of the information associated to them. In addition, a lack of consensus on the delimitation and characterisation of DMs has precluded re-usability of these costly resources.

Data-driven methods seem capable of motivating a DM classification at a low cost and partially avoiding the bias of human judges. Among the possible forms of characterising DMs, a hierarchical classification has the advantage of allowing for consistent procedures of information encoding, which can be automated and significantly reduce the cost of encoding information associated to DMs, while increasing its consistency.

The following section presents the basis of data-driven classification of DMs. After that, the data and clustering tool are described (Section 3). In Section 4, we present the classification obtained from a hand-tagged corpus, which we then extrapolate to automatically tagged corpus (Section 5). In Section 6, we discuss our results and sketch some future work, to finish with some conclusions.

* This research has been conducted thanks to a grant associated to the X-TRACT project, PB98-1226 of the Spanish Research Department. It has also been partially funded by projects HERMES (TIC2000-0335-C03-02) and PETRA (TIC2000-1735-C02-02).

[1] Other superficial clues, like punctuation or syntactic structures, can also mark discourse structure, but they are usually not considered DMs because they do not convey any information about the relations between discourse elements.

ISSN 1135-5948

2 Data-driven DM characterisation

The underlying hypothesis of data-driven DM characterisation is that *DMs with a similar behaviour in naturally occurring text will correspondingly have a similar behaviour as to the discourse processing instructions they elicit.* As follows, an automated classification of discourse markers according to features describing their occurrences in texts will mirror a taxonomy of the same items as discourse processing devices.

The main goal of clustering techniques is to identify partitions in an unstructured set of objects described by certain characteristics. Those partitions or classes contain similar objects according to some criteria, usually a distance or similarity function. Classes are expected to be different from each other, although sometimes they are not, since the method always produces classes, even if they don't exist in the domain. If the classes can be semantically interpreted by the human analyst, all the objects in a class can be considered together as a whole, and consequently treated in the same way.

Most of the previous work on obtaining data-driven DM characterisation relies on hand-coded examples (Siegel and McKeown, 1994; Litman, 1996; Di Eugenio, Moore, and Paolucci, 1997; Kim, Glass, and Evens, 2000). Common to the techniques of clustering and classification based on examples is their capacity of abstracting from a high number of examples and dealing with extensive sets of describing features. The main difference is that classification relies on preclassified examples. As said before, there is no standard on DMs to guide classification of examples, so any work based on a priori classifications of DMs will be controversial. In contrast, a hierarchical clustering algorithm can provide an objective, data-driven classification of DMs with an only source of bias, coming from the characterising features associated to the objects to be clustered.

Moreover, previous approaches to DM classification are oriented to knowledge requiring tasks, such as NLGeneration (Di Eugenio, Moore, and Paolucci, 1997; Kim, Glass, and Evens, 2000). For these tasks, the targeted characterisation of DMs has to be very rich, often beyond the capacities of existing NLP tools. That's why the learning examples are usually hand-coded, which supposes that the number of examples available for learning will be subordinated to the cost and capacities of tagging by human analysts.

However, a more shallow account of DMs can be useful for NLP tasks such as information retrieval (Corston-Oliver and Dolan, 1999), text summarisation (Ono, Sumita, and Miike, 1994; Marcu, 1997a) or even for obtaining a hierarchical structure of discourse (Marcu, 1997b). Such a shallow account can be obtained fully automatically, thus offering the portability and scalability needed for real-world applications. Trying to avoid the bottleneck of hand-coded examples, we have tried to obtain a satisfactory characterisation of DMs which can be applied with no need for manual tagging.

3 Experiment

3.1 Discourse Markers

We clustered occurrences of 577 Spanish DMs[2], including cue phrases and syntactical structures. These DMs were gathered from previous work on DMs for NLP (Marcu, 1997c; Knott, 1996), and specific approaches to DMs in Spanish: grammatical (Martín Zorraquino and Portolés, 1999), computational (the lexicon of the MACO morphological analyser for Spanish (Carmona et al., 1998)) and from a corpus study. They are described in an electronic lexicon by syntactic, discourse segmental and rhetorical information (see Table 1), oriented to obtaining a representation of discourse that is useful for automated text summarisation.

Each DM instance to be clustered was described by a set of 19 features, upon which the clustering tool evaluated similarity (see Table 2). The choice of features was motivated by previous research on classification of DMs (Siegel and McKeown, 1994; Litman, 1996; Di Eugenio, Moore, and Paolucci, 1997), which suggests that discourse structural features (level of embedding, segment markedness, surrounding words, ortography) are useful for describing DM behaviour. We additionally included features productive in the DM lexicon, like syntactical or rhetorical categories, taking care that they did not completely determine the classification.

A number of other possible features, such as verbal tense or argumental structure, were

[2]We worked with 784 expanded forms corresponding to 577 basic cue phrases

DM	boundary	syntactic type	rhetorical type	direction	content
además	not appl.	adverbial	satellizer	inclusion	reinforcement
a pesar de	strong	preposition	satellizer	right	concession
así que	weak	subordinating	chainer	right	consequence
dado que	weak	subordinating	satellizer	right	enablement

Table 1: Sample of the cue phrase lexicon

not considered for various reasons. In the first place, we restricted the number of describing features because they increased the complexity of the objects to be clustered and of the clustering solutions, which overloaded the clustering tool. Moreover, only those features were included that could be obtained automatically, either from the DM lexicon mentioned above or by shallow text processing tools. This permits to work with a high number of DM instances, which would have an unaffordable cost if the defining features had to be obtained manually.

3.2 Corpus

We extracted 68,275 random paragraph-sized occurrences of DMs from a corpus of 5.5 million words of balanced Spanish text (LEXESP) and 10.5 million of newspaper text.

To obtain some of the contextual features listed above, the corpus was morphosyntactically analysed (Carmona et al., 1998) and unambiguous intrasentential discursive segments and DMs were identified by an automated discourse segmenter (Alonso and Castellón, 2001).

All the used text processing tools priorise precision over recall. However, no disambiguation of DMs was performed, so that their relative degree of markedness could be reflected by their raw occurrences in text. As a consequence, the recall of the discourse segmenter was enhanced, so that all words that were formally identical to a DM in the lexicon were categorised as such, regardless of their ambiguity as to their discursive or sentential function. This increased recall implied a decrease in precision, with an error rate of 38%. Most of the mis-analysed words are relatives and coordinating conjunctions, very ambiguous and very frequent at the same time.

To assess the impact of this error rate, we obtained a random sample of the corpus containing 280 words which had been automatically categorised as DMs, and worked with three versions of it: in a *fully automatic corpus*, we worked with the whole of the 280 might-be-DMs, in the other two, only the 172 actual DMs were considered. From these two, the *automatic corpus* kept the automatic tagging information untouched and the *hand-tagged corpus* was manually revised. Classifications from hand-tagged corpus served as a comparison ground to interpret classifications from automatic corpora.

3.3 Clustering Tools

The software used to perform the cluster analysis of DMs is KLASS+ (Gibert, 1997), an autonomous clustering tool oriented to ill-structured domains. It applies an ascendant hierarchical method that builds classes iteratively clustering the most similar pair of objects at each step. The underlying cluster algorithm is chained reciprocal neighbours (De Rham, 1997), which is not order depending, so the final tree is always the same regardless of the objects ordering.

Similarity is calculated according to some distance measure or transformation. KLASS+ permits to work with different distances and similarity coefficients, including mixed distances that enable simultaneously working with categorical and numerical variables. In this application, χ^2 metric was used.

KLASS+ organises objects in a binary tree and the number of final clusters may be decided after the clustering. It offers some interpretation-oriented tools for helping in the analysis of the clustering results: using a heuristic criterion, it can recommend the best number of classes, it provides the prototypical description and the distribution of the variables for every class, either in numerical or graphical way, and it can identify *characteristic variables* (Gibert, Cortés, and Rodríguez-Roda, 2000; Gibert, Aluja, and Cortés, 1998), which can be used to identify the meaning of the final classes.

4 Establishing a Comparison Ground

The DMs described with the features obtained from the hand-tagged corpus were clustered and a hierarchical tree was found.

in-lexicon features	**possible values**
DM form	Form of the DM
Rhetorical content	RST-like values: Enablement, cause, circumstance, comparison, concession, etc.
Syntactical type	Adverb, anaphoric, coordinating, preposition, non-personal, subordinating
Rhetorical type	Connector, chainer, nucleizer, organizer, satellizer
contextual features	**possible values**
Occurrence in initial sentence	yes/no
Occurrence in final sentence	yes/no
Occurrence in initial segment	yes/no
Occurrence in final segment	yes/no
Position of DM in segment	Initial, middle, final
Previous word	Grammatical category of the following word
Following word	Grammatical category of the following word
Level of embedding	1, 2, 3, 4, 5, 6
Kind of segment of occurrence	Segments given by the discourse segmenter§
Kind of parent segment	Segments given by the discourse segmenter§ and sentence of occurrence
Kind of previous segment	Segments given by the discourse segmenter§
Kind of following segment	Segments given by the discourse segmenter§
Negation in the segment of occurrence	yes/no
Negation in the previous segment	yes/no
Negation in the following segment	yes/no

§ The segments identified by the discourse segmenter are: adjectival, adverbial, apposition, unmarked string, coordinated, marked weak, marked strong, non-personal, literal, marked, parenthetic, prepositional, relative

Table 2: Defining features of DMs for clustering

Using KLASS+ recommendations, a classification consisting of 3 classes was firstly obtained. The descriptive tools proposed by KLASS+ showed that negation and segment-contextual features perturbed classification, so they were left out of the objects' descriptions and the analysis was repeated.

4.1 Interesting Levels of Partition

At the 3-class level, classes are defined as:

1. **Extra-sentential** DMs are typically phrases outside the scope of a sentence carrying macro-structural discourse information, characterised by segment-middle position.
2. **Rightwards directed** DMs are grammatically integrated in the sentence, and syntactically linked to the segment at their right.
3. **Bi-directional** DMs are also grammatically integrated in the sentence, but they are equally attached to the text at each side of the DM.

As the number of classes increases, so does the number of characterising features, approaching the granularity of the descriptions by human analysts. An organisation of DM features can be mapped to a hierarchy of features ordered by their characterising power, mappable to a hierarchical classification of DMs. This hierarchy is now briefly presented.

4.2 Feature Relevance

Some of the features in the initial set, like *occurrence in initial/final segment* or *sentence* were not found to be characterising at any level of granularity, since they present very similar distribution in all the classes (see Figure 1, left). In contrast, characterising features, like *position of the DM in the segment*, present different values across classes, thus constituting a distinguishing variable of the class, which can be used to identify it either totally or partially (see Figure 1, right).

DM classes can be organised in a hierarchy that accounts for the characterising function of features at different levels of partition. This hierarchy is organised as follows:

1. At the topmost level, *position of the DM in the segment* distinguishes DMs occurring mainly in segment initial position from those in any other segment position.
2. DMs integrated in the sentence are further distinguished by *level of embedding*,

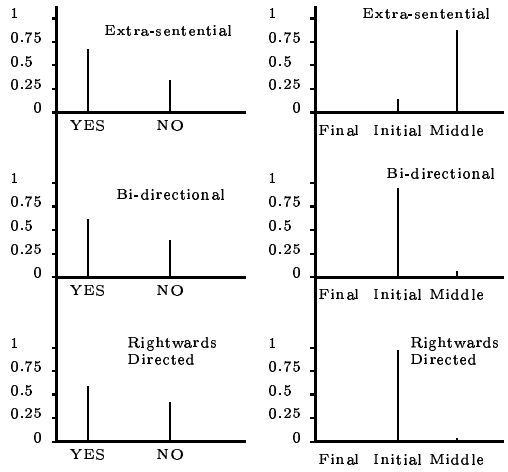

Figure 1: Distribution across classes for features *occurrence in initial sentence* (left) and *DM position in the segment* (right).

type of segment of occurrence and *type of parent segment*[3].

3. *Syntactical type* subdivides the rightwards directed class, often in correlation with and *rhetorical type.*
4. *rhetorical content* further differentiates classes.
5. At very particular levels, *form of DM* is a highly discriminating feature.

However, there is a group of DMs that escapes this feature hierarchy, which we have called *content-poor*, namely, relatives, some non-personal verb forms and, to a lesser degree, coordinating conjunctions. They can be considered unprototypical DMs because they lack strong characterising features, and, consequently, their position in a classification tends to fluctuate. Indeed, many DM theorists do not consider that relatives and non-personal verb forms perform a discursive function. However, we consider that they are indicative of fine-grained discourse structure that is useful for some NLP tasks, for example, text summarisation.

In contrast, prototypical DMs are well-established discourse operators with a defined discursive function. They are to be found in highly characterising contexts that explicitly signal their linguistic function. When they are clustered, they form stable classes, with a meaningful core of features.

[3]We detected a possible misrepresentation of these features, because they were found to be mutually redundant in most of the cases. Moreover, they are not characterising within extra-sentential DMs.

5 *Extrapolation from Hand-Tagged to Unsupervised Corpus*

As a result of errors in the automated tagging process, DMs in automatically tagged versions of the corpus have wrong values for some of the attributes. Moreover, the fully automatic corpus contains some words that are incorrectly categorised as DMs. To assess the impact of these errors in automatic clustering, the two sets of objects from the two automatic corpora were clustered, and the resulting classifications were compared to the ones obtained from the hand-tagged corpus.

The underspecification of content-poor, unprototypical DMs is found to be the main cause for the differences between hand-tagged and automatic classifications. Unprototypical DMs tend to perform a sentential function, so they are more grammaticalised and less marked in naturally occurring language. Consequently, there are few textual clues upon which a shallow NLP tool such as the discourse segmenter could disambiguate their discursive or sentential function. Since this tool priorises precision over recall, elements that cannot be safely described are left underspecified.

5.1 Comparison of Hand-Tagged vs. Automatic Corpus

Figure 2 presents the comparison between classifications from hand-tagged and automatic corpora. Class 2 in the automatic classification can be considered an equivalent to the class of right-directed, since it is mainly constituted by DMs classed as right-directed in the hand-tagged classification. In contrast, at first sight, classes 1 and 3 seem not to be comparable.

However, if we consider content-poor DMs as a factor of divergence between hand-tagged and automatic corpora, and in the corresponding classifications, Class 1 can be taken as an equivalent to the class of extra-sententials in hand-tagged classifications. Some bi-directionals, underspecified by the automated tagging process, have come to share some characterising features with extra-sententials, like middle segment position in unmarked segments. Therefore, they are grouped together in Class 1.

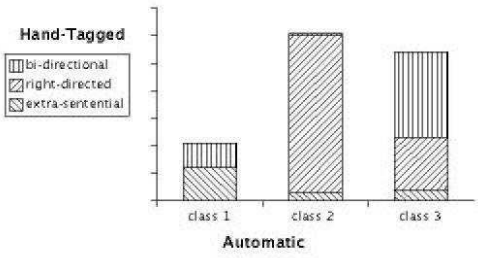

Figure 2: Difference of DM classes in classifications from hand-tagged and automatic corpus

Something similar happens to Class 3, which is mainly constituted by underspecified DMs. This class is best explained as grouping together content-poor DMs, very similar not only by rhetorical content, but also by their lack of characterising contextual or structural features. So, three meaningful classes can be identified in the automatic classification: **Extra-sententials** (Class 1), **Rightwards directed** (Class 2) and **Content-poor** (Class 3).

5.2 Comparisons with Fully Automatic Corpus

A third version of the corpus was clustered with all the might-be-DMs that had been detected automatically, including non-DMs, that is to say, instances of words which are formally identical to a DM but which are performing a sentential function. At first sight, this cluster analysis did not seem successful, because classes could not be interpreted by human analysts. The resulting classification presented some meaningless classes, and the feature hierarchy outlined in Section 4.2 was only partly recognisable.

We compared the classification from the fully automatic corpus to the two previous ones. In order to make classifications comparable, hand-tagged and automatic classifications were extended with an additional class containing all non-DMs in the corpus. The 8-class level of partition resulted the clearest to evidence correspondences between classifications.

As can be seen in Figure 3, classifications from hand-tagged and fully automatic corpora do not seem to be comparable, except for classes 6 and 7, which are mainly composed of right-directed DMs in both classifications.

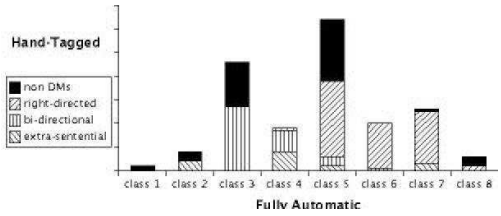

Figure 3: Comparison between hand-tagged and fully automatic classifications

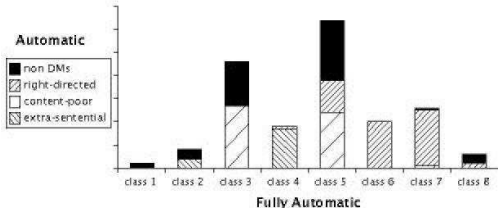

Figure 4: Comparison between automatic and fully automatic classifications

In contrast, classifications from automatic and fully automatic corpora show a higher degree of correspondence. In Figure 4, we can see that classes 6 and 7 can still be interpreted as composed of right-directed DMs. Moreover, classes 3 and 5 can now be interpreted in terms of DM prototypicality, as containing mostly unprototypical DMs, namely, non-DMs and content-poor DMs. In all classifications, extra-sententials (Class 4 in fully automatic) are grouped together with content-poor bi-directionals because, as we have said, the automatic tagging process produces an underspecification on their characterising features, which become very similar to those of extra-sententials.

We can see that, when a set of DMs and non-DMs is clustered, content-poor DMs are grouped together with non-DMs, thus evidencing that content-poor DMs are more similar to non-DMs than to prototypical DMs. However, they also show some similarity with prototypical DMs, as can be seen in Class 5 of Figure 4, mainly constituted by unprototypical DMs, but also containing some right-directed, prototypical DMs. Therefore, it can be argued that DM prototypicality is better explained gradually than as a binary distinction between DMs, with discursive function, and non-DMs, with sentential function, in contrast to the general practice in classification of DMs.

6 Discussion

The results of clustering are only useful if they can be semantically interpreted by the human analyst. However, while the classification obtained from a hand-tagged sample of the DM instances was meaningful, classifications of the same DM instances that were not manually revised were hard to interpret. Even when the organisation of DMs presented in Section 4 had been established, it was hard to recognise it in classifications of fully automatic corpora.

By comparing three classifications with varying degree of manual revision, the elements causing the differences between classifications could be identified. Differences were due to errors in the automatic tagging process and to the underspecification of DMs with low or none DM-prototypicality. When some of the causes of perturbance are removed, like in the automatic classification, where non-DMs are excluded, the organisation of DMs is much more comparable to the one in Section 4. Once the perturbing elements have been isolated, classifications of fully automatic corpora can be re-interpreted, and corpus samples with no manually revised correlate can be successfully analysed.

Preliminary experiments with other corpus samples show that the organisation of classes and features described in Section 4 can be recognised across classifications from fully automatic corpus. Consistent classes can be found at the 6- or 8-class level of partition. Further work includes an extensive analysis of fully automatic corpus, applying resampling techniques to make bootstrap-oriented clustering.

The comparison of different kinds of classifications has evidenced semantic relations between DM features that can help in characterising them more adequately. For example, in automatic classifications, extra-sententials are grouped together with content-poor bi-directionals because the latter are underspecified by the automated tagging process. However, some segment-structural features (*kind of segment*, *level of embedding*) could help in discriminating them, providing a better account of their organisation. However, the similarity metric used for clustering relies exclusively on syntactic information, and it fails to capture such semantic relations. Nonetheless, semantic information can be incorporated to clustering by *clustering based on rules* (Gibert, Aluja, and Cortés, 1998). This method, which KLASS+ implements, allows finding the structure of an ill-structured dataset by introducing a semantic bias. Further work includes implementing the semantic relations discovered so far by rules.

7 Conclusions

Our work shows the utility of clustering as a portable and scalable technique for discourse processing technologies. We have shown that DMs can be satisfactorily described by features obtained from a fully automatic process. Comparison of classifications with varying degree of manual revision adequately assesses the impact of the error rate of the automatic tagging process, so that meaningful DM classes that can be identified across samples.

The classifications found by KLASS+ have clearly delimited DM classes defined by a stable core of features with varying degree of specificity, which can also be expressed in terms of a hierarchy of features. This hierarchy can ground a classification of DMs that enables the use of inheritance for reducing the cost of encoding information associated to them, while guaranteeing consistency. It has also proven useful in a lexical acquisition approach to DMs, to build a tool for DM extraction (Alonso, Castellón, and Padró, 2002).

We have proposed the concept of gradual prototypicality as an alternative to the classic binary distinction between sentential and discursive function of DMs (Hirschberg and Litman, 1993), by comparing classes across different kinds of classifications. We have found that prototypical DMs show a stable behaviour across classifications, and are classed according to a meaningful set of features, whereas unprototypical DMs fluctuate across classifications and are classed together with non-DMs, if there are any. This empirical ground can be useful to delimit the concept of DM.

References

Alonso, Laura and Irene Castellón. 2001. Towards a delimitation of discursive segment for natural language processing applications. In *First International Workshop on Semantics, Pragmatics and Rhetoric*, Donostia - San Sebastián, November.

Alonso, Laura, Irene Castellón, and Lluís Padró. 2002. X-tractor: A tool for extracting discourse markers. In *LREC 2002 workshop on Linguistic Knowledge Acquisition and Representation: Bootstrapping Annotated Language Data.*

Carmona, Josep, Sergi Cervell, Lluís Màrquez, M. Antònia Martí, Lluís Padró, Roberto Placer, Horacio Rodríguez, Mariona Taulé, and Jordi Turmo. 1998. An environment for morphosyntactic processing of unrestricted spanish text. In *First International Conference on Language Resources and Evaluation (LREC'98)*, Granada, Spain.

Corston-Oliver, Simon H. and W. Dolan. 1999. Less is more: Eliminating index terms from subordinate clauses. In *37th Annual Meeting of the Association for Computational Linguistics (ACL'99)*, pages 348 – 356.

De Rham, C. 1997. La classif. hierarch. selon la méthode des voisins réciproques. *Cahiers d'Analyse des Données*, V(2):135–144.

Di Eugenio, Barbara, Johanna D. Moore, and Massimo Paolucci. 1997. Learning features that predict cue usage. In *ACL-EACL97, Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 80–87, Madrid, Spain.

Gibert, Karina. 1997. The use of symbolic information in automation of statistical treatment of ill-structured domains. *Artificial Intelligence Communications.*

Gibert, Karina, Tomàs Aluja, and Ulises Cortés. 1998. Knowledge discovery with clustering based on rules. interpreting results. In *Principles of Data Mining and Knowledge Discovery.* Springer-Verlag.

Gibert, Karina, Ulises Cortés, and Ignasi Rodríguez-Roda. 2000. Identifying characteristic situations in wastewater treatment plants. In *Workshop in Binding Environmental Sciences and Artificial Intelligence.*

Hirschberg, Julia and Diane Litman. 1993. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, 19(3):501–529.

Kim, Jung Hee, Michael Glass, and Martha W. Evens. 2000. Learning use of discourse markers in tutorial dialogue for an intelligent tutoring system. In *COGSCI 2000, Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*, Philadelphia, PA.

Knott, Alistair. 1996. *A Data-Driven Methodology for Motivating a Set of Coherence Relations.* Ph.D. thesis, University of Edinburgh, Edinburgh.

Litman, Diane J. 1996. Cue phrase classification using machine learning. *Journal of Artificial Intelligence Research*, 5:53–94.

Marcu, Daniel. 1997a. From discourse structures to text summaries. In Mani and Maybury, editors, *Advances in Automatic Text Summarization*, pages 82 – 88.

Marcu, Daniel. 1997b. The rhetorical parsing of natural language texts. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 96 – 103, Madrid, Spain, 7 - 12 July.

Marcu, Daniel. 1997c. *The Rhetorical Parsing, Summarization and Generation of Natural Language Texts.* Ph.D. thesis, Department of Computer Science, University of Toronto, Toronto, Canada.

Martín Zorraquino, M. Antonia and José Portolés. 1999. Los marcadores del discurso. In Ignacio Bosque and Violeta Demonte, editors, *Gramática Descriptiva de la Lengua Española*, volume III. Espasa Calpe, Madrid, pages 4051–4213.

Ono, K., K. Sumita, and S. Miike. 1994. Abstract generation based on rhetorical structure extraction. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING-94)*, pages 344 – 348, Kyoto, Japan.

Siegel, Eric V. and Kathleen R. McKeown. 1994. Emergent linguistic rules from inducing decision trees: Disambiguating discourse clue words. In *AAAI94, Proceedings of the 12th Conference of the American Association for Artificial Intelligence*, pages 820–826.