An Analytic Account of Discourse Markers for Shallow NLP

Laura Alonso*, Jennafer Shih*, Irene Castellón*, Lluís Padró†

*Departament de Lingüística General
Universitat de Barcelona

†Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya

Abstract

We present a feature-based approach to the description of discourse markers (dms) oriented to automated discourse analysis for shallow Natural Language Processing (NLP). Dm describing features have been chosen based on previous work, descriptive adequacy and our concrete NLP needs and capacities. An organization of these features in dimensions of dm meaning has been inferred via data-driven techniques, and finally implemented in a computational dm lexicon.

1 Introduction

The aim of this paper is to present a description of discourse markers (*dms*) to facilitate their exploitaition in automatic discourse analysis with shallow Natural Language Processing (NLP) techniques. Our final target is to obtain a representation of discourse to improve information condensation applications, like sentence compression or automatic text summarization. Therefore, we focus on the properties of *dms* as indicators of relative relevance and coherence relations between discourse units.

There are a number of mechanisms in language that allow establishing or eliciting coherence relations between parts of text, like certain kinds of pauses and prosodic contours, the choice of syntactic structures or lexic, and also a some lexical items which we call "discourse markers" (*dms*). What is most frequently understood under *dms* are spoken language items like *you know* or *well*, (Schourup 1985; Schiffrin 1987), but this concept can also include written language items like *but* or *because*.

Written *dms* have been specially studied in NLP as evidence of linguistic relations beyond the clausal core, the so-called *coherence relations*. Knott (1996) exploits this evidence to propose a data-driven typology of coherence relations. The opposite direction has also been explored: *dms* have been exploited as systematic signals of one or more of a pre-established set of relations, to identify or generate coherence relations in text. For example, Marcu (2000) and Corston-Oliver (1998) associate *dms* to the relations proposed by the Rhetorical Structure

Theory (Mann and Thompson 1988) in order to obtain a tree-like representation of discourse that allows to identify coherence and relevance relations in text. Schilder (2002) associates *dms* to the relations proposed by the Segmented Discourse Representation Theory (Asher 1993; Lascarides and Asher 1993).

As follows, the utility of *dms* for systematic, eventually automated discourse analysis lies in the assumption that they can be consistently associated with one of a set of coherence relations, thus providing rich information about the structure of texts by very simple NLP techniques, like pattern matching and lexicon lookup. This makes them crucial for discourse analysis based on shallow NLP techniques, as is our case.

However, the association of *dms* to relations is not exempt of problems, because *dms* are highly polysemous, both with respect to their sentential or discursive function (Hirschberg and Litman 1993) and with respect to the relation they convey (Jayez and Rossari 1998; Di Eugenio et al. 1997, among others). Polysemy is even more problematic when only shallow NLP techniques can be used, because no alternative sources of evidence[1] are available to support a decision as to the possible relation indicated by an ambiguous *dm*.

In this paper we propose a description of the discursive relations signalled by *dms* aimed to minimize this polysemy problem. Our approach supposes a drastic reduction on the informativity of *dms*, but increases their tractability and reliablity within a shallow NLP framework.

In the first place, we reduce the number of decisions to be made when assigning meaning to a *dm*, that is, we reduce the set of relations with which *dms* can be associated. We propose a set of very basic, coarse-grained discursive effects that meet our descriptive necessities and NLP capabilities.

In the second place, we underspecify those aspects of the meaning of *dms* that are more ambiguous. To allow underspecification of only part of the meaning of *dms*, we take a compositional approach to describing their semantics. In other words, we don't associate *dms* to atomic relations, but rather to a conglomerate of the proposed basic meanings, so that *dms* can be reliably characterized in those aspects in which they are not ambiguous. Then, we attempt to discover the organization of these features of *dm* semantics in order to systematize them.

The rest of the paper is structured as follows: the next Section exposes briefly our working delimitation of *dms*. In Section 3, the compositional approach to the semantics of *dms* is compared with relation-based approaches, and the set of features to describe *dm* semantics is presented. In Section 4, the distribution of these features is analyzed via data-driven techniques, and we propose an organization of features in three different dimensions of discursive meaning. Section 5 describes how this organization has been implemented in a computational *dm* lexicon. We finish with some conclusions and future work.

2 Working delimitation of *dms*

The concept of "discourse marker" is a controversial one. We will not attempt a definition in this paper, we will just expose a working delimitation whereupon our study on *dm* semantics will build.

A starting set of 80 *dms* was created to study their properties in depth. This initial set was created in parallel for English, Spanish and Catalan, so that *dms* were only included if they

[1] For example, no reference resolution, no propositional representation, not even full syntactic parsing in most cases.

had a near-synonym in the three languages. We considered cross-linguality as an indication that the meaning conveyed by the *dm* is indeed a basic meaning. We believe that, after these basic meanings have been established, they will provide a reference framework to carry out further, language-specific distinctions as needed. Besides cross-linguality, *dms* were included in this initial set for the following reasons:

- they illustrate a basic discursive meaning that is useful for the identification of relative relevance and coherence relations in texts (*besides*, *in fact*).
- they are treatable by shallow NLP techniques, like pattern matching to access databases. As a result, only lexical items have been considered.
- they have been extensively studied in previous work (Knott 1996; Marcu 1997; Martín Zorraquino and Portolés 1999) (*however*, *presently*).
- they are frequent in written, formal corpora (*for example*, *after*), which allows to obtain huge amounts of examples to apply data-driven methods.

3 A feature-based account of *dms*

3.1 *Dm* semantics as a set of features

Dms have been often described by defining a set of processing instructions, also called discourse or coherence relations, and associating each *dm* to one or more of them. However, a consenus set of those processing instructions has never been established, on the contrary, a major problem for theories of discourse is the confusing proliferation of coherence relations (Hovy and Maier 1992). Moreover, *dm* description relies very much on subjective judgements, and are often highly biased by application needs or description requirements.

Feature-based descriptions try to overcome some of the subjectivity in the description of *dms* by decomposing their complex meaning of into simpler components. These components of meaning can be treated as primitives in the processing of discourse (like *cause*), or else their effects are systematically characterized by textual evidence.

An explicitly analytic approach provides a framework that supports partial descriptions of *dm* semantics, thus allowing different granularities in the analysis, in other words, naturally enabling underspecification or specification mechanisms. Moreover, a transparent description of *dm* semantics facilitates portability of lexical resources, in this case, a *dm* lexicon, to different frameworks and applications.

Sanders, Spooren, and Noordman (1992) propose four cognitive primitives to account for the coherence effects of discourse relations:

(i) *basic operation*: the relation between discourse entities is causal or additive,
(ii) *polarity*: the relation is negative or positive,
(iii) *source of coherence*: the relation has a semantic or pragmatic interpretation
(iv) *order*: in causal relations, the segments can be presented in their basic order or the reverse.

Building on Sanders et al. (1992), and after analyzing several relation-based theories of text coherence, Knott (1996) presents a set of relations to account for discourse coherence mostly aimed at NLGeneration. These relations are based on the behaviour of *dms* in text, which is described with a set of 8 features:

- source of coherence: semantic or pragmatic
- anchor: cause-driven or result-driven
- pattern of instantiation: unilateral or bilateral
- focus of polarity: anchor-based or counterpart-based
- polarity: positive or negative
- presuppositionality: presupposed or non-presupposed
- modal status: actual or hypothetical
- rule type: causal or inductive

As Knott, we take a data-driven approach, but our direction of research is the opposite: we do not exploit *dms* to establish a set of coherence relations, but try to exploit the discursive effects of coherence relations to characterize *dms*. Another difference between our work and previous feature-based work is the orientation to shallow NLP. Accordingly, we disregard features that resort to the deep understanding of texts, but focus on those for which stable correlates can be found in a representation of texts obtained via shallow NLP techniques.

3.2 Features of *dm* semantics for shallow NLP

Taking into account previous work and application objectives (relevance and coherence assessment of discourse units) and restrictions (shallow NLP), a set of eight features were chosen to describe the semantics of *dms*. A very strong criterion for the choice of features was the fact that they could be treated either as semantic primitives or else associated to objective tests that relied on systematic features that characterise their context of occurrence, mostly discursive mechanisms like ortographic correlates of pauses (punctuation marks), modality markers, negation or uneven distributions of lexic or referential expressions.

In what follows we provide an intuitive description of the semantics of each of these features (summarized in Table 1). *Dms* are in italics, segments dominated by the *dm* are underlined.

context describes the setting for a discourse entity, as in Langacker (1987)'s *figure-ground*
`People started demonstrating` *as soon as* `the war began`.
discursive effect: *A discourse segment holding a context relation reduces the amount of contextual inferences from the discourse segment to which it is attached, by specifying its context.*

This feature was chosen because it contributes to determine the lack of relevance of the dominated segment. Moreover, this relation is very frequent in texts, so it provides an important amount of information for NLP applications.

causality elicits a causal relation between elements, as in Kehler (2002) or Asher and Lascarides (2003)'s *cause*.
`They lost the elections` *because* `they manipulated information`.
processing primitive: *causal relation between elements.*

Cause is a basic feature of textual coherence (Halliday and Hasan 1976; Sanders, Spooren, and Noordman 1992).

parallelism makes relevant an equivalence relation between elements, of the type *A is (mod) B*, where *mod* is optional and can be any kind of modifier (*kind of*, *contrary of*, etc.) (Hobbs 1985; Kehler 2002)

feature	discursive effect	dimension
context	provides the setting for a discourse entity *People started demonstrating as soon as the war began.*	matter
parallelism	establishes an equivalence between two elements *A is (mod) B* *Some other governments supported the war, as in Spain.*	matter
causality	elicits a causal relation between two elements *They lost the elections because they manipulated information.*	matter
revision	negates some previous information, explicit or implied *No weapons of mass destruction were found, but Iraq was invaded.*	matter
progression	introduces a new topic or intention *The "Prestige" wandered about for a week, and it finally sunk.*	argumentative
elaboration	continues a presented topic or intention *The "Prestige" wandered about for a week, all along the coast.*	argumentative
symmetric	attachement to a node at the same level in the discourse tree *They lied to voters and so they lost the elections.*	structure
asymmetric	attachment to a node in a different level in the discourse tree *Because they lied to voters, they lost the elections.*	structure

Table 1: Proposed features to describe the semantics of *dms*. The rightmost column displays the dimension of discursive meaning to which each feature has been assigned, as discussed in Section 4.

`Some other governments supported the war,` *as* `in Spain.`
discursive effect: *There are various discourse effects that are often associated to parallelism, for example: the elements involved display syntactically parallel constructions, there is a presence of equating operators (and, or), or the elements belong to a common semantic type and this type is relevant.*

Parallelism is a very productive feature of relations between elements at all levels of language. In discourse, it is a basic feature of coherence between segments, and can contribute to select an adequate semantic interpretation for the entities which are equated, by transitivity of their features.

revision negates some previous information, explicit or implied, the kind of negation can range from topic-based contrast (Umbach to appear) to denial of expectation, going through semantic oppostion and counterargument (Lagerwerf 1998).
`No weapons of mass destruction were found,` *but* `Iraq was invaded.`
processing primitive: *implicatures and the fact that they are defeasible.*

Revision is a highly marked mechanism in language, usually exploited to attract the attention of hearers. Therefore, it signals the relevance of the segments involved, as well as a strong coherence relation between them.

elaboration provides further information on an already presented topic or continues an already stated intention (Grosz and Sidner 1986).
`The "Prestige" wandered about for a week,` *all along* `the coast.`
discursive effect: *Topic and referential continuity.*

progression introduces a new discourse topic or intention (Grosz and Sidner 1986).
`The "Prestige" wandered about for a week, and it` *finally* `sunk.`

discursive effect: *Topic and referential discontinuity.*

Both progression and elaboration describe the contribution of a discourse segment to the main topic of a text, so they are very useful to determine its relevance.

symmetric in a tree-like structure of discourse, a discourse unit equivalent to a node is attached to another node at the same level (Polanyi 1988; Webber 1988)
`They lied to voters` *and so* `they lost the elections.`
discursive effect: *Coordinability (Gómez Txurruka 2000)*[2]

asymmetric in a tree-like structure of discourse, a discourse unit equivalent to a node is attached to another node at a different level (Polanyi 1988; Webber 1988)
Because `they lied to voters, they lost the elections.`
discursive effect: *Sentential subordination, interchangeability of discourse units.*

It can be assumed that relative position in a hierarchical structure correlates with relevance. We do not consider *asymmetric* equivalent to SDRT *subordinating* (Asher and Vieu 2001), because *asymmetric* is a purely syntactic feature, while discursive subordination involves many other kinds of discursive meaning. *Symmetric* and *asymmetric* are actually the two sides of a same relation, but they were distinguished in order to perform the data-driven analysis presented in Section 4.

We have not considered as a semantic property of discourse relations how they may constrain the possible attachment point of discourse units, or the number and kinds of arguments they may take. Indeed these aspects of *dms* are crucial for their integration in automatic discourse parsers (Marcu 2000; Forbes et al. 2003), but we consider this aspect of the meaning of *dms* is qualitatively different to the ones proposed so far, and beyond the scope of this work.

We argue that these features allow to make relevant distinctions, as can be seen in Figure 1, where three different kinds of cause usually distinguished in relation-based descriptions of *dms* can be very well distinguished.

Contrastively, the distincitions that are not made are either not possible in our framework or are not relevant for our representation purposes. For example, we do not make a distinction between semantic and pragmatic cause (Sweetser 1990) because this is beyond the capacities of shallow NLP, and they have the same effects for the assignment of relevance and coherence to discourse segments.

Another important factor in the configuration of this set of features was the fact that all the features were productive. This means that all features were distinctive in most of the combinations of features, that is to say, that none was redundant with another. We believe that this is based in the fact that these features belong to different dimensions of discursive meaning. In the following section, we try to provide empirical support for the distribution of features in dimensions.

4 Empirical analysis of *dm* semantics

In order to analyze the organization of the proposed features, they were applied to the description of the semantics of our starting set of prototypical *dms* , in order to build an initial

[2]Gómez Txurruka (2000) argues that, if an "and" can be used to link two Discourse Representation Structures, these are related by a *coordinating* relation, equivalent to *symmetric*.

consequence → cause + progression + coordinating (*so that*)

> They sold the best diamonds *so that* by the time the eastern goods reached the Mediterranean only the most uninspiring of the diamonds were left.

reason → cause + elaboration + subordinating (*because*)

> The judge is believed to have added 25 to each sentence specifically *because* the police had carried out the attack while operating in their official capacity.

purpose → cause + progression + subordinating (*in order to*)

> *In order to* fully appreciate Distant Voices, Still Lives, we would have to be familiar with the work of Lowry, Lawrence, Larkin.

Figure 1: Distinctions that can be made within the causal family, by combination of the proposed features of *dm* semantics.

Description of *dms* in the lexicon: *however*		Description of relations between discourse segments in corpus: ... button sewing, a laborious work that involves, *nonetheless*, a certain tremendism.	
context	no	context	no
causality	underspecified	causality	no
parallellism	underspecified	parallellism	no
revision	**yes**	revision	**yes**
progression	**yes**	progression	**yes**
elaboration	no	elaboration	no
symmetric	**yes**	symmetric	**yes**
asymmetric	no	asymmetric	no

Figure 2: Illustration of how correlation coefficient is obtained from lexicon and corpus descriptions. The correlation coefficient between features assigned a positive value increases (*revision*, *progression*, *symmetric*), as it also increases between features assigned a negative value (*context*, *elaboration*, *asymmetric*), while it decreases between features assigned a different value (for example, *progression* and *symmetric*).

dm lexicon. Additionally, they were also used in corpus annotation. Then, the distribution and relations of features were studied.

In the *dm* lexicon, each *dm* was manually assigned a positive or negative value for each of the eight features, according to its core discursive effects in extensive journalistic corpora for Catalan, Spanish and English. In corpus annotation, three human judges determined whether each of the features applied for the relation between each discourse segment and the discourse segment where it was attached to, independently of the presence of a *dm*. In both cases, the presence of features was identified by systematic tests exploiting their effects on text.

Then, the patterns of co-occurrence of describing features were obtained by calculating the correlation coefficient between them. The correlation was calculated between co-occurrences of features in the same element, be it a description of a *dm* in the lexicon or a relation between discourse segments in corpus annotation, as can be seen in Figure 2.

The correlation coefficient ranges from 0 to 1, where correlation coefficients close to 0 are not informative of feature organization, but negative correlation values elicit complementary

distribution of features, meaning that they belong to the same paradigm. It is trivial that *symmetric* and *asymmetric* present values close to -1, because, as said before, these two features are the two sides of the same discursive meaning. However, keeping them separated allows to test the consistency of the proposed method and also to establish relations between the rest of features more clearly.

As for *progression* and *elaboration*, they also present a correlation coefficient close to -1, but their correlation is higher in the description of corpus than in the description of lexicon. This is mainly due to the fact that these features are mostly realized in text by coherence mechanisms other than *dms* , like thematic or referential cohesion. Therefore, most *dms* are underspecified with respect to these features, whereas they are almost always recognizable in text, where all coherence mechanisms are available for interpretation.

The relation between *causality*, *context* and *parallelism* is not so clear. In the *dm* lexicon they present correlations strong enough to be thought to belong to the same paradigm (.6, .4, .3), but this strength is weakened in the descriptive text (.5, 2., 1.), and data from the argumentative text allow no conclusions to be drawn at all (.2, .1, .1). However, agreement between judges only decreases from $kappa = .6$ in the descriptive text to $kappa = .5$ in the argumentative one, indicating that such weak correlations are proper of text organization and not a result of disagreement between judges. Finally, *revision* never presents a strong correlation coefficient with any of the other features.

Drawing from the organization of features elicited by their correlations, and resorting to semantic criteria based on previous work and corpus analysis, features were clustered in three dimensions of *dm* meaning, as summarized in Table 1:

structure (*symmetric, asymmetric*) accounts for the hierarchical organization of discourse in a tree-like structure like that proposed by Polanyi (1988) or Webber (1988).

argumentation (*progression, elaboration*) describes general mechanisms of discourse planning *à la* Grosz and Sidner (Grosz and Sidner 1986).

matter (*revision, causality, parallelism, context*) accounts for extra-argumental relations between the propositional content of discourse segments. It is comparable to Halliday and Hasan (1976)'s *internal*, Mann and Thompson (1988)'s *subject-matter*, or Kehler (2002)'s discourse relations.

	caus	**paral**	**revision**	**elab**	**prog**	**coord**	**subord**
context	-.34	**-.60**	-.02	.34	-.33	-.20	.18
causality	-	**-.40**	-.07	-.17	.20	.07	-.06
parallelism	-	-	.12	-.14	.10	.13	-.12
revision	-	-	-	-.03	-.05	.05	-.05
elaboration	-	-	-	-	**-.78**	.05	-.05
progression	-	-	-	-	-	.43	**-.42**
symmetric	-	-	-	-	-	-	**-.99**

Table 2: Confusion matrix of the correlation coefficient between the distribution of *dm* semantic features in the description of *dms*.

The data obtained from the distribution of *revision* in the description of lexicon and corpus does not allow to drive any safe conclusions about its relations with other features. However,

	caus	paral	revision	elab	prog	coord	subord
context	-.15	**-.47**	-.04	.47	-.43	-.22	.27
causality	-	-.21	-.06	-.14	.15	-.15	.17
parallelism	-	-	-.05	-.39	.44	.52	-.47
revision	-	-	-	-.09	.06	.07	-.09
elaboration	-	-	-	-	**-.92**	-.52	.56
progression	-	-	-	-	-	.57	-.52
symmetric	-	-	-	-	-	-	**-.94**

Table 3: Confusion matrix of the correlation coefficient between the distribution of *dm* semantic features in corpus annotation of an eminently descriptive text (average kappa agreement between annotators $kappa = .64$).

	caus	paral	revision	elab	prog	coord	subord
context	.13	-.15	-.07	.17	-.19	-.16	.18
causality	-	-.19	.07	-.14	.16	.08	-.01
parallelism	-	-	-.16	.10	-.02	.25	-.19
revision	-	-	-	-.26	.32	.32	-.29
elaboration	-	-	-	-	**-.85**	-.49	.54
progression	-	-	-	-	-	.54	-.46
symmetric	-	-	-	-	-	-	**-.89**

Table 4: Confusion matrix of the correlation coefficient between the distribution of *dm* semantic features in corpus annotation of a highly argumentative text (average kappa agreement between annotators $kappa = .54$).

it can be argued that *revision* operates on the propositional content of discourse segments, considering that implicatures are obtained from propositional content in a systematic way. Moreover, its combination with features in the *argumentative* dimension allow to distinguish well-established discourse relations, like *concession* (revision + elaboration, *although*) or *contrast* (revision + progression, *but*).

5 Implementation of *dm* semantics in a computational lexicon

Once the organization of *dm* semantics has been established, it has been implemented in a starting computational *dm* lexicon.

Semantic ambiguity of *dms* is treated by underspecification at the level of dimensions of meaning, with differences across languages. For example, Catalan *perquè* (*because*) is ambiguous in the argumentative dimension, expressing either progression or elaboration, while its parallels in Spanish (*porque*) and English (*because*) prototypically signal elaboration.

The flexibility of the feature-based description of *dms* allows effortless adaptations to different domains or needs of representation of discourse, since relevant features of *dms* can be easily isolated and used for every task. It constitutes a basic knowledge base for a variety of NLP applications working at discourse level, for example, e-mail summarization (Alonso et al. 2003).

6 Conclusions and Future Work

We have presented a feature-based account of discourse markers oriented to shallow NLP applications. Features have been proposed taking into account previous work, application needs and linguistic evidence from corpus. A data-driven analysis of the distribution of these features has been carried out, and they have been organized in three dimensions of *dm* semantics: *argumentation*, *matter* and *structure*.

The proposed description has been implemented in a small computational lexicon in three languages. Usual problems in using *dms* in computational applications, like ambiguity, have been adressed by exploiting the organization of the semantic features of *dms* in meaning dimensions. This lexicon is currently being used for automated *dm* acquisition and for shallow discourse analysis, more concretely, to obtain a partial structure of discourse that allows for relevance and coherence assessment in text summarization.

7 Acknowledgements

We would like to thank Ezequiel Andújar Hinojosa and Roberto Sola Salvetierra for lengthy discussions on problematic annotation issues. Also Henk Zeevat provided extensive comments on preliminary versions of this work, as well as audiences at the ESSLLI'03 workshop *Discourse Particles: Theory and Implementation* and at the Research Workshop of the Department of General Linguistics at the Universitat de Barcelona. We would also like to thank two anonymous reviewers for substantial comments on an earlier version of this paper. This research has been conducted thanks to the grant PB98-1226 of the Spanish Research Department and project PETRA (TIC-2000-1735-C02-02).

References

Alonso, L., B. Casas, I. Castellón, S. Climent, and L. Padró (2003). Combining heterogeneous knowledge sources in e-mail summarization. In Recent Advances in Natural Language Processing (RANLP 2003), Borovets, Bulgaria.

Asher, N. (1993). Reference to abstract objects in discourse. Kluwer Academic Publishers.

Asher, N. and A. Lascarides (2003). Logics of Conversation. Cambridge University Press.

Asher, N. and L. Vieu (2001). Subordinating and coordinating discourse relations. In Intl. Wkshp. on Semantic, Pragmatics and Rhetorics, San Sebastián.

Corston-Oliver, S. H. (1998). Computing representations of the structure of written discourse. Ph. D. thesis, University of California, Santa Barbara.

Di Eugenio, B., J. D. Moore, and M. Paolucci (1997). Learning features that predict cue usage. In ACL-EACL97, Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, Madrid, Spain, pp. 80–87.

Forbes, K., E. Miltsakaki, R. Prasad, A. Sarkar, A. Joshi, and B. Webber (2003). D-LTAG system - discourse parsing with a lexicalized tree-adjoining grammar. Journal of Language, Logic and Information.

Gómez Txurruka, I. (2000). The semantics of 'and' in discourse. Technical Report ILCLI-00-LIC-9, ILCLI, University of the Basque Country.

Grosz, B. and C. Sidner (1986). Attention, intentions, and the structure of discourse. Computational Linguistics 3(12), 175–204.

Halliday, M. A. K. and R. Hasan (1976). Cohesion in English. English Language Series. Longman Group Ltd.

Hirschberg, J. and D. Litman (1993). Empirical studies on the disambiguation of cue phrases. Computational Linguistics 19(3), 501–529.

Hobbs, J. (1985). On the coherence and structure of discourse. Technical Report CSLI–85–37, Center for the Study of Language and Information, Stanford University, Calif., USA.

Hovy, E. and E. Maier (1992). Parsimonious or profligate: How many and which discourse structure relations? Technical Report RR-93-373, NTIS No. ADA 278715, USC Information Sciences Institute.

Jayez, J. and C. Rossari (1998). Discourse relations vs discourse marker relations. In ACL'98 Workshop on Discourse Relations and Discourse Markers, pp. 72–78.

Kehler, A. (2002). Coherence, Reference, and the Theory of Grammar. CSLI Publications.

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

Lagerwerf, L. (1998). Causal Connectives Have Presuppositions; Effects on Coherence and Discourse Structure. Den Haag, Holland Academic Graphics.

Langacker, R. W. (1987). Foundations of Cognitive Grammar, Volume I: Theoretical Prerequisites. Stanford: University Press.

Lascarides, A. and N. Asher (1993). Temporal interpretation, discourse relations and commonsense entailment. Linguistics and Philosophy 16(5), 437–493.

Mann, W. C. and S. A. Thompson (1988). Rhetorical structure theory: Toward a functional theory of text organisation. Text 3(8), 234–281.

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

Marcu, D. (2000). The rhetorical parsing of unrestricted texts: A surface-based approach. Computational Linguistics 26(3), 395–448.

Martín Zorraquino, M. and J. Portolés (1999). Los marcadores del discurso. In I. Bosque and V. Demonte (Eds.), Gramática Descriptiva de la Lengua Española, Volume III, pp. 4051–4213. Madrid: Espasa Calpe.

Polanyi, L. (1988). A formal model of the structure of discourse. Journal of Pragmatics 12, 601–638.

Sanders, T. J. M., W. P. M. Spooren, and L. G. M. Noordman (1992). Toward a taxonomy of coherence relations. Discourse Processes 15, 1–35.

Schiffrin, D. (1987). Discourse Markers. Cambridge University Press.

Schilder, F. (2002). Robust discourse parsing via discourse markers, topicality and position. Natural Language Engineering 8(2&3). special issue on robust methods in analysis of natural language data.

Schourup, L. C. (1985). Common Discourse Particles in English Conversation. New York: Garland.

Sweetser, E. (1990). From Etymology to Pragmatics. Metaphorical and Cultural Aspects of Semantic Structure. Cambridge University Press.

Umbach, C. (to appear). Contrast and information structure: A focus-based analysis of 'but'. Linguistics.

Webber, B. L. (1988, June 7-10,). Discourse deixis: Reference to discourse segments. In Proceedings of the 26th Annual Meeting of the ACL, State University of New York at Buffalo, pp. 113–122.