

A framework for feature-based description of low level discourse

Laura Alonso Alemany and Ezequiel Andújar Hinojosa and Robert Sola Salvatierra
GRIAL

Departament de Lingüística General
Universitat de Barcelona

Abstract

We present the preliminary stage of the discursive annotation of a corpus in Spanish. In contrast with previous discourse annotation projects, we have applied a feature-based description of relations between spans of text, and we have focussed in the low level organization of discourse.

This first phase has served to settle the annotation framework and to provide empirical support to theoretical claims, like a hierarchy of markedness in the meaning of discursive relations. Systematic guidelines have been established for the interpretation of relations between discourse units, so that tests and interpretation procedures have been associated to determine the presence of the features characterizing discourse relations.

1 Introduction and Motivation

Annotated corpora have shown to be very useful to build, evaluate or improve linguistic resources and computational analyzers. Big corpora with annotation at sentential and sub-sentential level (morphologic, syntactic, semantic) are widely available, but they are scarce for discourse and, what is more preoccupying, they are quite fragmentary, because of differing theoretical frameworks and heterogeneous phenomena that are annotated.

Nevertheless, more and more efforts have been devoted to developing discursively annotated corpora in recent years. Various annotation schemata have been developed in order to improve the consistency of annotation and to reduce the annotation cost (Discourse Resource Initiative, 1997; Cooper et al., 1999; Carlson and Marcu, 2001).

In this paper we present a schema for discursive annotation of Spanish written corpora focussed in the low level organization of discourse and that applies a feature-based (vs. relation-based) approach to the description of relations between discourse units.

This annotation is specially oriented to improve high-level NLP applications, like text summarization and question answering, for Spanish. Therefore, we have focussed in those aspects of discourse that allow to identify coherence relations and relative relevance, and we have disregarded those distinctions that are beyond the capabilities of the state-of-the-art analyzers for our target languages.

We focus in the **low level of discourse** because it constitutes a natural enhancement of the NLP pipeline towards discourse, but also because a reliable and informative representation of low level discourse can be obtained with shallow linguistic information, which comes short to resolve ambiguities at higher levels.

A **feature-based** annotation schema seems to be descriptively more adequate than relation-based approaches, like those based in the Rhetorical Structure Theory (RST) (Mann and Thompson, 1988). In these approaches, relations between discourse units are described by assigning them an atomic label out of a pre-determined set of possible relations. These atomic labels fail to capture the complexity of the meaning of discourse relations. In some cases, heterogeneous information (for example, intentional and subject-matter) is implicitly mixed in a single label, and so it is not possible to treat each kind of information separately. In other cases, such heterogeneous information is encoded in disjunct labels, and so relations can only be characterized by one of them, while it has often been argued that discourse relations can convey various kinds of meaning simultaneously.

In contrast, if relations between discourse units are described compositionally, heterogeneous meanings can be expressed in a distinct, transparent manner. Such a description is arguably more adaptable to different necessities of representation, and it is also comparable with other theoretical frameworks. Moreover, the interpretation of discourse relations can be systematized by associating tests and decision procedures to each distinct feature, improving the consistency of annotation.

The rest of the paper is structured as follows. Our theoretical framework is sketched in Section 2, and Section 3 describes a schema of corpus annotation based in this framework. The results of an exploratory corpus annotation are discussed in Section 4, and some implications are put forward.

2 The low level of discourse

The low level of discourse is constituted by those phenomena that cannot be explained within a clausal scope (like anaphora, coherence relations, illocutionary force and the like) but that are mostly based in the features of utterances at clausal level (like syntactic constructions, referential expressions or discourse particles, among others).

Our description of the low level of discourse is focussed on the coherence relations between minimal discourse units. More concretely, we aim to describe extra-argumental relations between sentence constituents that have the status of discourse unit, and also relations between sentences that are comparable to these. We believe that these relations are basic linguistic mechanisms that hold for virtually every kind of text, even for dialog, and can be described independently of higher-level discourse devices, like genres or domains.

2.1 Minimal discourse units

For space constraints, we are not providing extensive grounding here on the nature of discourse units. In general, our position is based in the classical LDM, as recently synthesized by Polanyi et al. (2004).

Two kinds of minimal discourse units are considered: *discourse segments*, minimal units of content at discourse level, and *discourse markers*, functional units that convey information on the relations between segments. However, the distinction between discourse segment and marker is not clear cut. In some cases, discourse segments can function as discourse markers, as in example (1), where the underlined segment elicits a relation between its matrix clause and a previous discourse unit, in this case, a grouping of segments. This is why our

annotation schema allows to characterize discourse segments with the functional information that is characteristic of discourse markers.

- (1) No satisfecho con todo ello, el ciudadano vasco Urquiza desoía a su partido y cada día se tomaba sus potes y vermuts en el "batzoki" del PNV de Durango.
Not happy with all this, the basque citizen Urquiza didn't listen to his party and he had his "potes" and vermuths in the "batzoki" of the PNV in Durango.

Unlike Carlson and Marcu (2001), we consider clausal subjects and objects of verbs and collocations of saying and opinion as discourse segments, because it has been shown that their status in discourse is equal to that of a stand-alone segment, while their matrix clause performs a function comparable to that of lexical discourse markers (Verhagen, 2001), as illustrated in the following examples, where the construction with a verb of opinion (2) conveys the same discursive meaning as that with a discourse marker (3).

- (2) [...] estoy convencido de que dentro de poco superará el examen de plancha, [...]
[...] I am sure that she will pass the ironing exam in short, [...]
- (3) [...] seguramente dentro de poco superará el examen de plancha, [...]
[...] she will surely pass the ironing exam in short, [...]

Distinguishing subclausal segments yields a finer-grained representation of discourse. Such representation is descriptively more adequate, and will correspondingly improve the performance of NLP applications like question answering or extreme summarization, for which require a very precise representation of text. But even for applications that require less precision, like Information Retrieval, sub-sentential analysis has been shown to improve results (Corston-Oliver and Dolan, 1999).

2.2 A feature-based description of discourse relations

We propose that discourse relations are described compositionally, as a conglomerate of basic components of meaning, and not by assigning them an atomic label, as is the case in most work on discourse annotation. Atomic labels provide no good account of the complexity of discourse relations, because they do not allow for heterogeneous information to be expressed simultaneously and distinctly.

We have established eight basic features that describe the meaning of discourse relations in various dimensions of discourse meaning simultaneously, summarized in Table 1. These features do not aim to provide a complete description of relations between discourse segments, instead, we have focussed in those aspects of discourse that are useful for text summarization and question answering, mainly those that allow to identify coherence relations between discourse segment and also their relevance in text. Moreover, we have not considered finer-grained distinctions that are beyond shallow NLP capabilities, such as the distinction between different kinds of cause or between different sources of coherence (e.g., semantic or pragmatic).

We have distinguished three co-existing dimensions of discursive meaning (Alonso et al., 2003):

matter (*context, parallelism, cause, revision*) conveying the kind of meaning of Mann and Thompson (1988)’s subject-matter, but at the level of granularity of Kehler (2002),

argumentative (*progression, elaboration*) equivalent to Grosz and Sidner (1986)’s attentional / intentional level, and

structural (*symmetric, asymmetric*) expressing the relation between a discourse segment and the node where it is attached in a tree-like structure of discourse as that proposed by Webber (1978) or Polanyi (1988).

Within each of these dimensions, the range of possible meanings is ordered in a scale of markedness, from the default representation of discourse for that dimension to the most marked one.

Structural features are not equivalent to SDRT *coordinating* and *subordinating* (Asher and Vieu, 2001), because *asymmetric* does not always imply *subordinating*, as in example (2), where the underlined segment holds an asymmetric relation with its matrix clause but it is not discursively subordinated. In our annotation schema, discursive subordination is obtained compositionally: a subordinating relation is an asymmetric relation of elaboration. For some representation purposes, interactions with some matter meaning may determine whether such a relation is considered subordinating or not, for example, it can be stipulated that causal relations are never subordinating, independently of their structural and argumentative features.

Our feature-based account is specially adequate to deal with ambiguity, both for human and automated analyses. In relation-based approaches, ambiguity is resolved in a single decision with a

	argumentative		
	matter	structural	
result	cause	prog	
consequence	cause	prog	sym
purpose	cause	prog	asym
cause	cause	elab	
reason	cause	elab	asym
sequence	parallel	prog	sym
summary	parallel	prog	asym
restatement	parallel	elab	
example	parallel	elab	asym
contrast	revision	prog	sym
counterargument	revision	prog	asym
concession	revision	elab	asym
narration	context	prog	sym
conclusion	context	prog	asym
background	context	elab	
circumstance	context	elab	asym

Table 2: Equivalence between clusters of features and usual relation-like labels. Note that *narration* is the result of combining the least marked features, coherently with the role of this relation in Kehler (2002) and Asher and Lascarides (2003).

big range of possibilities, and the only possibility to deal with unclear cases is to leave the relation fully underspecified, that is, to provide no information at all for that relation. In a feature-based approach, meaning is partitioned and so are decisions, so that underspecification can affect only part of the meaning of the relation, and not the whole of it. The complexity of meaning is also distributed, so that there is a smaller range of choices for each decision, which makes decisions easier to take and then more reliable.

A feature-based account is specially adequate to integrate heterogeneous kinds of information in a systematic way. For example, discourse markers are very informative of matter and structural dimensions, but most of them do not convey strong argumentative information. Consequently, relations between segments marked by a discourse marker may obtain their argumentative information from the topical or intentional structures of text, or from the syntactical structures of the utterance, while the rest of features can be determined by the discourse marker.

Lastly, this feature-based description is compatible with relation-based ones, as seen in Table 2.

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

Table 1: The proposed basic components of meaning of discourse relations, organized in three dimensions. The rightmost column displays the markedness of features within each dimension (- - means least marked, ++ means most marked).

3 Annotation of Spanish text

We have applied the presented framework to the annotation of a small journalistic corpus in Spanish. The purpose of this application was twofold: first, to test the descriptive adequacy of the proposed components of meaning, then, to settle an annotation schema for the annotation of a bigger corpus.

Three judges have taken part in this preliminary annotation, two naive judges, with no background in linguistics, and a linguist. The purpose of having naive judges annotate the corpus was to check the intuitivity of the proposed features. Judges made decisions basically relying on their intuitions, but, since the task is not very natural in itself, an annotation manual was created to direct decisions. This manual was enhanced and refined during annotation, and is now ready for use in annotation at a big scale (Alonso et al., 2004).

The corpus to be annotated was of journalistic genre, with some variation in subgenres: television critic, everyday stories, and opinion. We chose a varied corpus in order to check the applicability of features in different kinds of text. The annotation was carried out in XML, an example can be seen in Figure 1. The annotated corpus can be found at <http://lingua.fil.ub.es/~lalonso/discor/>.

The procedure for annotation was as follows: once the text had been read completely, discourse segments were characterized one by one in order of occurrence in the text, by the following features:

node of attachment identification number of the discourse unit to which a *segment* is at-

```

<seg id="1.2" type="sentence">
  <attach="1.7" sym="" asym="1"
  cause="1" context="" parallel=""
  revision="" elabora="1" progre=""
  gl_cause="because"
  gl_elabora="he has been killed">

  <seg id="1.2.1" type="segment">
    <attach="1.2" sym="" asym="1"
    cause="" context="" parallel=""
    revision="1" elabora="1" progre=""
    gl_revision="if you have such surnames,
    then you are a good basque"
    gl_elabora="Urquiza couldn't be
    a good basque">

    <seg id="1.2.1.1" type="disc_mark">
      <attach="1.2,1.2.1" sym="" asym="1"
      cause="" context="" parallel=""
      revision="1" elabora="1" progre=""
      gl_revision="if you have such surnames,
      then you are a good basque"
      gl_elabora="Urquiza couldn't be
      a good basque">

      ____Despite____
    </seg>

    ____his____
    ____surnames____
  </seg>
  ...
</seg>

```

Figure 1: Example of XML annotation of a sentence (1.2), a sub-sentential discourse segment (1.2.1) and a discourse marker (1.2.1.1).

tached, or identification numbers of the segments related by a *discourse marker*¹,
features the presence or absence of each of the eight proposed features, by application of decision trees that reflect the theoretical characterization of features, as seen in Figure 2,
glosses if a *matter* or *argumentative* feature was present, a gloss was provided for it, as an elicitation protocol aimed to improve consistency of annotation and also to make finer-grained distinctions of features.

The values assigned for **node of attachment** are not totally determined by the kind of discourse unit. Segments with a relating function (as in example (1)) are assigned the same kind of value as a discourse marker, that is, the identification numbers of the segments they relate. Contrastively, discourse markers that do not relate discourse segments, like those relating the text with the author (for example *luckily* or *of course*), can be attached to the discourse segment where they are found.

Each segment was related to at least one other segment. When there was no clear attachment node or when the segment was the starting point or main claim of a text, it was considered the top of its local structure of discourse, and it was attached to itself.

Each segment carried a tag for its own attachment to other segments, but none for the attachment of other segments to it. The decision as to which segments carried the attachment tags was based on their markedness: the most marked segment in a relation was the one that carried the tag, usually, segments under the scope of a discourse marker or characterized by redundant lexic or referential expressions². In case the criteria of markedness was not applicable, segments coming later in discourse were the ones to carry the tags.

Glosses were aimed to guarantee consistency and keep track of decision procedures, but they also served to encode finer-grained distinctions within each of the proposed features. These distinctions have not considered basic in our framework because they are beyond our NLP capabilities, but they are widely accepted in the literature and we believe it is interesting to study their realization in corpus. The following glosses were provided:

¹Available attachment points are not only minimal discourse segments, but also discourse units constituted by continuous spans of discourse segments

²Note that markedness has no relation with RST's nuclearity, since marked segments can be the nucleus or the satellite of a relation.

revision elicit the information that is denied, implicitly or explicitly, in four different forms, in order of markedness, the first three based in Lagerwerf (1998), the least marked based in Umbach (to appear):

denial of expectation elicit expectation
concessive elicit *tertium comparationis*
opposition the related segments can be rephrased with a correction *but (sino)*
focus-based if none of the others apply

cause one of four discourse markers with which the relation can be paraphrased: *in order to* (purpose), *because* (cause), *that's why* (reason), and *therefore* (consequence)

parallelism the common class of things to which the two segments belong, which has to be salient in the context, either because it is lexicalized or by regular abstraction procedures, like hyperonymy.

Additionally, glosses for progression and elaboration were also provided, summarizing the topic or intention that was being introduced or elaborated, respectively. This procedure increased the consistency of the annotation for these two features.

When judges were unsure about a feature for a relation, they left it underspecified. This resulted in an average of 2.6 features per relation, which increases slightly (2.7) for segments containing a discourse marker and decreases to 2.2 for relations at the beginning of a paragraph.

Some issues were specially difficult for annotators to reach an agreement, like the annotation of discontinuous discourse markers (*so... that...*) or an effective procedure to systematize the assignment of matter features to syntactical constructions like relative clauses or absolute participles. They have been given provisional solutions, but will be closely studied in further annotation.

3.1 Consistency of preliminary annotation

The corpus consisted of 6 articles totalling 3541 words (from 207 to 1042 words) in 154 sentences. Texts had been previously segmented in discourse minimal units (discourse segments and discourse markers). 468 intrasentential segments were found, of which 84 were argumentally required (as in example 2) and the rest were different kinds of adjuncts, of which 226 were dominated by a discourse marker. 261 discourse markers were identified, corresponding to 101 different forms, the most frequent are *y* (*and*, 50 occurrences), *para* (*for*, *to*, 20 occurrences), *como* (*like*, *as*, 15 occurrences) and *pero* (*but*, 9 occurrences).

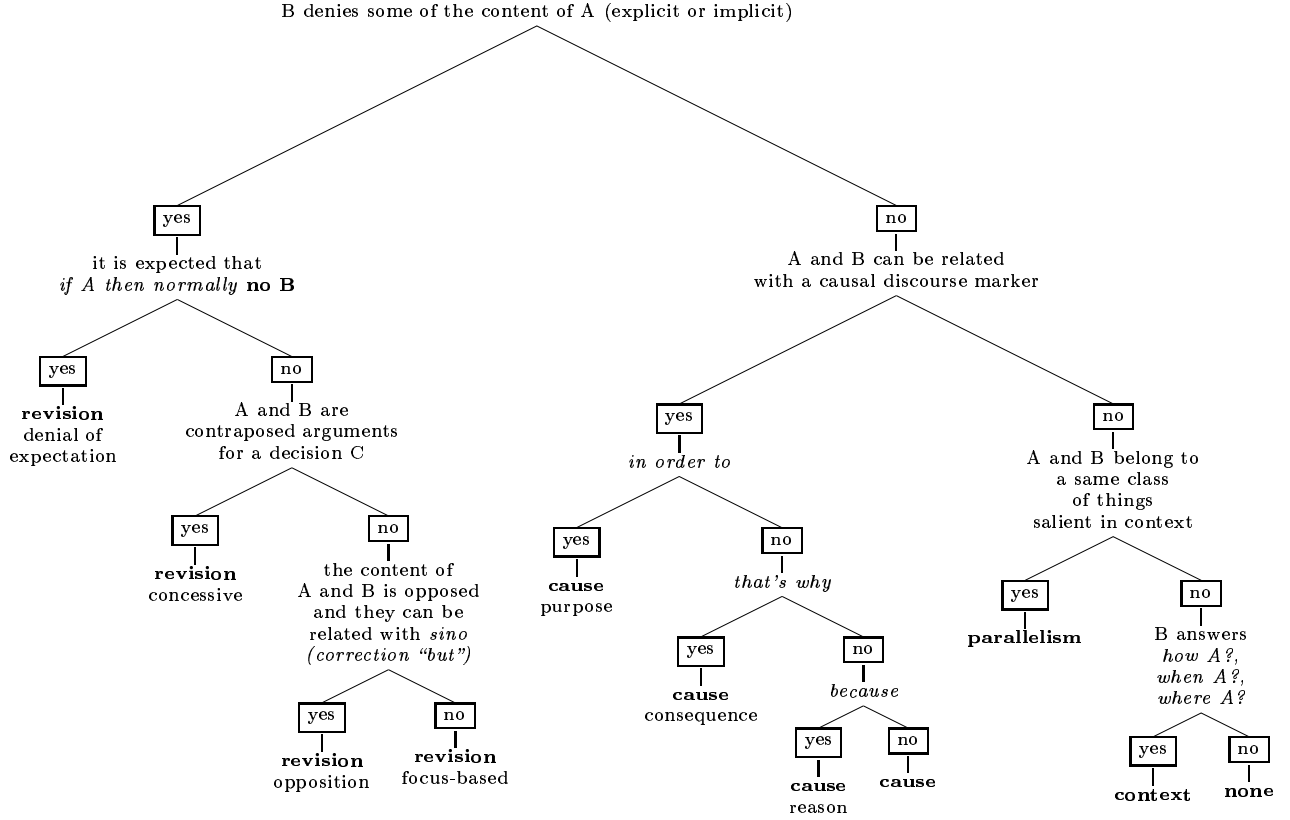


Figure 2: Decision tree for annotating features belonging to the matter dimension. It reflects the characterization of features presented in Table 1: decisions are taken in order of markedness, so that it is first discriminated whether the most marked feature (*revision*) is present or not. Finer-grained distinctions within *revision* and *cause* are also directed by the decision tree.

	units	features	revision	cause	parallel	context	prog.	elab.	sym.	asym.	average
text 1	181	531	0.55	0.69	0.75	0.32	0.64	0.59	0.52	0.49	0.57
text 2	74	194	0.59	0.17	0.56	0.47	0.42	0.46	0.54	0.46	0.46
text 3	261	716	0.74	0.64	0.51	0.36	0.54	0.51	0.60	0.58	0.56
text 4	49	131	0.89	0.51	0.50	0.61	0.52	0.48	0.71	0.72	0.62
text 5	153	405	0.57	0.53	0.49	0.38	0.39	0.35	0.42	0.48	0.45
text 6	163	437	0.85	0.74	0.59	0.42	0.56	0.52	0.51	0.52	0.59
total	881	2414	0.70	0.55	0.57	0.43	0.51	0.48	0.55	0.54	0.54
discourse markers			0.72	0.53	0.53	0.39	0.44	0.43	0.39	0.42	0.48
beginning paragraph			0.03	-0.0	0.16	0.01	0.29	0.18	0.03	0.11	0.10

Table 3: Average kappa agreement between judges for the six texts used in the preliminary stage of corpus annotation.

The consistency of the annotation was evaluated by kappa agreement between the values of the features assigned to each node. As can be seen in Table 3, the average agreement, $\kappa = .54$ is quite low, and does not guarantee a good reproducibility of the results (Carletta, 1996). However, it has to be taken into account that this agreement has been obtained in the preliminary phase of annotation, during the process when annotation criteria were being established, and also that the judges were not professional annotators. It has to be noted, even with trained professional annotators, Carlson et al. (2003) present $\kappa = .6$ in the initial stages of annotation, reaching $\kappa = .75$ at the end of the project.

4 Discussion of the results of annotation

Interestingly, kappa agreement for discourse markers is significantly lower than the average, except in the case of the *revision* feature, which presents the highest agreement per feature. However, the number of features for characterizing discourse markers was higher than the average, which means that judges recognized them as highly informative of discourse organization but were not able to distinguish clearly the components of their meaning. Their characterization will have to be studied more closely.

As can be seen by the low agreement for segments occurring at the beginning of paragraph, they were very difficult to characterize for judges, which seems to indicate that the coherence mechanisms that apply at paragraph level are qualitatively different from those at sentential and inter-sentential level. This seems a strong argument to treat the low level organization of discourse as an autonomous level of representation of language.

The fact that *revision* presents the highest agreement of all features seems to provide empirical support to the hierarchy of markedness proposed in our framework.

Although judges had been allowed to relate segments to more than one discourse unit by more than one relation, this never happened in the corpus. This supports the claim that discourse can be represented as a hierarchical tree. In our framework, the structure of discourse is described as a superposition of the hierarchical trees obtained from the representations of each dimension. In other words: segments do not have more than one relation in each dimension, although they might be attached to different segments in different dimensions, as illustrated in Figure 3.

5 Conclusions and Future Work

We have presented the application of a feature-based approach to the description of the low level discourse organization to corpus annotation. The annotation of a small corpus in Spanish has been carried out, and an annotation framework has been established in the form of systematic procedures for decision taking. These procedures are a direct mapping of the theoretical characterization of the basic components of meaning of discourse relations.

The consistency of this preliminary annotation is comparable to that of the preliminary stages of other annotation initiatives. The fact that two of the three annotators involved had no background in linguistics provides support for the validity of the proposed features as basic components of the meaning of discourse relations. Moreover, the analysis of consistency results seems to validate the markedness hierarchy proposed for features within dimensions of meaning.

Future work will be aimed to the application of this annotation framework to a Spanish corpus of a reasonable size, most probably a subset of the CLiC-TALP corpus, which is currently being annotated with syntactic functions. When this corpus is annotated, there will be enough data to draw statistically significant conclusions on the relations between basic components of meaning within and across dimensions, and also to characterize sub-kinds of relations that are currently disregarded in our framework but described in annotation.

6 Acknowledgements

This research has been partially funded by MCyT program - BFF2001-5440 and also thanks to grant PB98-1226 of the Spanish Research Department.

References

- Laura Alonso, Jennafer Shih, Irene Castellón, and Lluís Padró. 2003. An analytic account of discourse markers for shallow nlp. In Manfred Stede and Henk Zeevat, editors, The Meaning and Implementation of Discourse Particles, workshop at ESSLLI'03.
- Laura Alonso, Daniel Alonso, Ezequiel Andújar and Robert Sola. 2004. Anotación discursiva de corpus en español. Technical Report GRIAL report series 2004-2, Universitat de Barcelona.
- Nicholas Asher and Alex Lascarides. 2003. Logics of Conversation. Cambridge University Press.
- Nicholas Asher and Laure Vieu. 2001. Subordinating and coordinating discourse relations. In Intl. Wkshp. on Semantic, Pragmatics and Rhetorics, San Sebastián.

- a. John loves Barolo.
- b. So he ordered three cases of the '97.
- c. But he had to cancel the order
- d. because *then* he discovered he was broke.

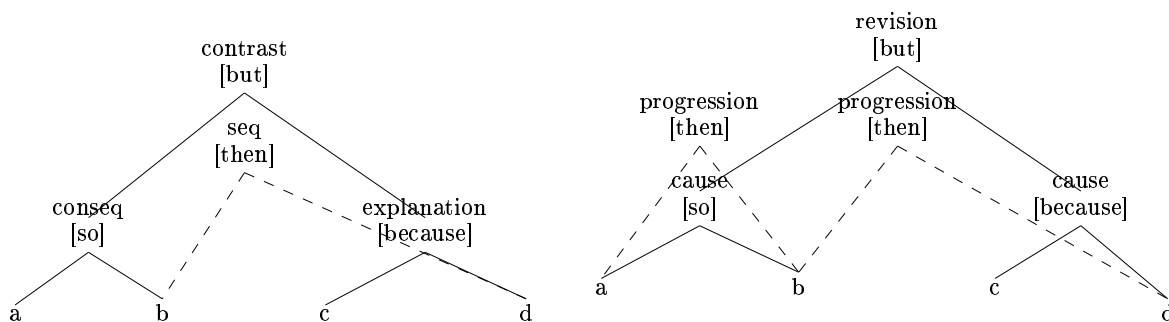


Figure 3: The example in the left is taken from Webber et al. (2003, pp. 551). If a single dimension of discourse is considered, this example can only be well described by a DAG. (2003) avoid to treat this kind of cases as a DAG by exploiting the anaphoric properties of averbial discourse markers like *then*. We explain that by resorting to different dimensions of discursive meaning, as seen in the example in the right, where solid lines represent the matter dimension, dashed lines represent the argumentative dimension and the structural dimension is represented by the relative level of terminal discourse units.

- Jean Carletta. 1996. Assessing agreement on classification tasks: the kappa statistic. Computational Linguistics, 22(2):249–254.
- Lynn Carlson and Daniel Marcu. 2001. Discourse tagging manual. Technical Report ISI-TR-545, ISI.
- Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2003. Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In Jan van Kuppevelt and Ronnie Smith, editors, Current Directions in Discourse and Dialogue. Kluwer Academic Publishers.
- Robin Cooper, Staffan Larsson, Colin Matheson, Massimo Poesio, and David Traum. 1999. Coding instructional dialogue for information states. Technical Report D1.1, The TRINDI Consortium.
- Simon H. Corston-Oliver 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.
- Discourse Resource Initiative. 1997. Standards for dialogue coding in natural language processing. Technical Report 167, Dagstuhl-Seminar.
- Barbara J. Grosz and Candace L. Sidner. 1986. Attention, Intention, and the Structure of Discourse. Computational Linguistics, 12(3):175–204.
- Andrew Kehler. 2002. Coherence, Reference, and the Theory of Grammar. CSLI Publications.
- Luuk Lagerwerf. 1998. Causal Connectives Have Presuppositions; Effects on Coherence and Discourse Structure. Den Haag, Holland Academic Graphics.
- William C. Mann and Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organisation. Text, 3(8):234–281.
- Livia Polanyi, Chris Culy, Martin van den Berg, Gian Lorenzo Thione, and David Ahn. 2004. A rule based approach to discourse parsing. In SIGDIAL 2004.
- Livia Polanyi. 1988. A formal model of the structure of discourse. Journal of Pragmatics, 12:601–638.
- Carla Umbach. to appear. contrast and information structure: A focus-based analysis of 'but'. Linguistics.
- Arie Verhagen. 2001. Subordination and discourse segmentation revisited, or: Why matrix clauses may be more dependent than complements. In Ted J. M. Sanders, Joost Schilperoord, and Wilbert P. M. Spooren, editors, Text Representation. Linguistic and psychological aspects, pages 337–357. John Benjamins.
- Bonnie Lynn Webber, Matthew Stone, Aravind K. Joshi, and Alistair Knott. 2003. Anaphora and discourse structure. Computational Linguistics.
- Bonnie Lynn Webber. 1978. A Formal Approach to Discourse Anaphora. Ph.D. thesis, Harvard University. Reprinted in Outstanding Dissertations in Linguistics Series, Garland Publishers, 1979.