

Measuring the Strength of Linguistic Cues for Discourse Relations

Fatemeh Torabi Asr and Vera Demberg

Cluster of Excellence Multimodal Computing and Interaction (MMCI)

Saarland University

Campus C7.4, 66123 Saarbrücken, Germany

`fatemeh@coli.uni-saarland.de`, `vera@coli.uni-saarland.de`

Abstract

Discourse relations in the recent literature are typically classified as either explicit (e.g., when a discourse connective like “because” is present) or implicit. This binary treatment of implicitness is advantageous for simplifying the explanation of many phenomena in discourse processing. On the other hand, linguists do not yet agree as to what types of textual particles contribute to revealing the relation between any pair of sentences or clauses in a text. At one extreme, we can claim that every single word in either of the sentences involved can play a role in shaping a discourse relation. In this work, we propose a measure to quantify how good a cue a certain textual element is for a specific discourse relation, i.e., a measure of the *strength* of discourse markers. We will illustrate how this measure becomes important both for modeling discourse relation construction as well as developing automatic tools for identifying discourse relations.

Keywords: Discourse relations, Discourse markers, Discourse cues, Implicitness, Implicit and explicit relations.

1 Introduction

Clauses, sentences and larger segments of a text should be connected to one another for a text to be coherent. A connection in the semantic-pragmatic level is established with the help of sharing entities in the discourse or relations between statements, which are called *discourse relations*. The discourse relations are usually described in terms of their relation sense (e.g., causal, temporal, additive). Identification of these relations, i.e., first coming up with a set of possible relation senses and then assigning labels to the segments of a given text, is an essential first step in both theoretical and application-based studies on discourse processing. Given a set of sense labels (like the ones in the Penn Discourse Treebank (PDTB, Prasad et al., 2008)), identification of the relations between neighboring segments of a text is a difficult task when the text segments do not include an explicit discourse connector. For example, in (1-a) the connective “because” is a marker of a causal relation between the two clauses, whereas in (1-b) the relation is not marked explicitly with a discourse connector.

- (1) a. Bill took his daughter to the hospital, because she looked pale and sick in the morning.
- b. I was very tired last night. I went to sleep earlier than usual.

The presence of explicit cues makes it easier for humans to infer discourse relations during comprehension of a text or an utterance. Similarly, explicit discourse connectors have been shown to help the automatic identification of these relations for NLP tools (Pitler et al., 2008). In fact, choosing a set of relation types in preparing discourse-level annotated corpora is often done with reference to the well-known lexical or phrasal discourse markers¹. A good example is the procedure used by the annotators of the Penn Discourse Treebank (PDTB) to identify implicit relations in the corpus². Some relations are associated with discourse cues which mark them almost unambiguously (e.g., “because” for a causal relation), while other discourse relations typically occur with no explicit marker (e.g., list relations), or tend to be expressed using markers which are ambiguous (e.g., synchronous temporal relations are usually marked by “while”, which can also be a cue for juxtaposition). One can look at this ambiguity from a different perspective: some discourse markers such as “but” are used in almost every type of adversative context, whereas a marker such as “unless” is used only for a very specific type of relation (disjunctive).

In this paper, we try to elaborate on the two-way link between discourse markers and the relation senses that are typically used in the literature. We propose a quantification of the cue *strength*, i.e., how well a discourse marker makes a discourse relation explicit in the text. Based on the numbers we extract from the PDTB, we suggest that implicitness should be treated as a continuum and not as a binary feature of a discourse relation. The rest of this paper is organized as follows: Section 2 introduces the probabilistic measure we use to estimate the strength of a discourse cue in marking a particular discourse relation between segments of a text. Section 3 includes a brief introduction of the PDTB hierarchy of relation

¹We use the terms *discourse marker* and *discourse cue* interchangeably in this paper. Nevertheless, *cue* is used more typically when the predictive nature of the textual element is highlighted (see Müller (2005) for a discussion on the terminology).

²As the case study reported in this paper has been done on the PDTB, we adapt their terminology when referring to different types of discourse cues and senses of discourse relations.

senses, statistics about distribution of implicit and explicit relations, and specifically, the strength measurements we performed on the annotated discourse connectives in the corpus. In the last section we discuss why and how consideration of the cue strength would help theoretical and application-based studies of discourse processing.

2 Quantification of the Marking Strength

A discourse relation is established between two (or more) segments of a text each of which includes several words or phrases. Applying a formal logic approach (like the one by Sanders et al. (1992)) would suggest that discourse relation establishment is an operation which takes place between independent arguments (statements) by means of explicit operators (discourse cues) or the relational semantics we obtain implicitly from the text according to our world knowledge. Although all words in the arguments contribute to the shaping of the relation, discourse cues as defined in the literature typically refer to a specific category of words or phrases which have an operator-like function in the discourse level. For example, Stede (2011) distinguishes discourse *connectives* as closed-class, non-inflectable words or word groups syntactically from adverbial, subordinate/coordinate conjunction, or preposition categories which themselves can only be interpreted successfully when they appear in a relation between two discourse segments. (Prasad et al., 2010), however, suggest that a variety of expressions exist that mark discourse relations, but they are not from the typically-considered syntactic categories, and in some places they are not even structurally frozen (e.g., “that would follow”).

Whatever syntactic or semantic function a discourse cue is associated with, the relative frequency of its occurrence in a particular type of discourse relation is what makes it interesting. Our focus is not on the structural properties of a discourse marker, but instead on the strength of the marker for indicating a specific discourse relation. Given a segment of a text, perhaps composed of two sentences whose discourse relation is to be determined, one would think about a set of cues that express the polarity and temporality of the sentences, the stated relation between the involved entities, as well as the presence of any word or expression that can be attributed to a specific discourse relation. A simple probabilistic model would look for a relation r which maximizes $p(r|cues)$. For estimating the probability of a discourse relation r given a cue cue , we can use Bayes’ theorem to formulate:

$$p(r|cue) = \frac{p(cue|r)}{p(cue)} * p(r) \quad (1)$$

where $p(r)$ is the prior probability of relation r , and $\frac{p(cue|r)}{p(cue)}$ determines the effect of the present discourse cue in identification of r , namely, the *strength* of the cue. If a word or expression is a good marker for a particular relation, we would expect it to have a high strength value. It would mean that the cue is seen in many instances of that relation relative to its total number of occurrences. We propose that the strength of a discourse marker is a reliable measure one can use to estimate how well that cue would work in a discourse relation identification task, be it by human comprehenders, annotators or computational automated tools.

3 Case study: PDTB

The Penn Discourse Treebank (Prasad et al., 2008) includes annotations of 18,459 explicit and 16,053 implicit discourse relations in texts from the Wall Street Journal. Explicit

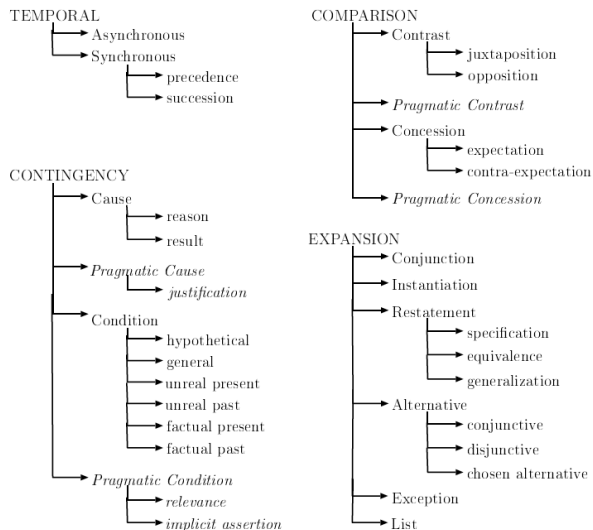


Figure 1: Hierarchy of senses in PDTB (Prasad et al., 2008)

relations are those expressed through one out of a closed-class set of discourse connective in the original text. After the annotation of explicit discourse connectors, annotators were asked to decide on the discourse relationship between any two adjacent sentences in the corpus which were not already linked through explicit connectors, and insert one or more suitable discourse connectives. Labeling of the relations is done according to a hierarchy of senses (see Figure 1), including four top-level classes: **CONTINGENCY**, **COMPARISON**, **TEMPORAL** and **EXPANSION**. In most of the cases the relation sense is chosen from the deepest possible level of the hierarchy (leaves of the tree). But when the annotators did not agree on the fine-grained relation sense (e.g., **Instantiation**), they compromised by tagging the relation with a more general sense (**EXPANSION** in this case).

In our study of cue strength, we decided to analyze only those relations for which the most specific tagging was available, i.e., those tagged with one of the 30 relation senses in the leaves of the hierarchy. In this set of relations we found 95 connective types which appeared in the explicit relations and 70 connective types used for annotation of the implicit relations. Strength values reported in this paper are calculated according to the explicit occurrences of a particular connective for a particular relation sense in the mentioned subset of text extracted from the PDTB. The strength values range between 0.0028 and 71.4370 after applying simple add-1 smoothing to avoid division by zero³.

3.1 Implicit vs. Explicit Relations

First of all, by looking at the overall distribution of relation types, we found a significant difference between implicit and explicit occurrences. Some types of relations tend to appear implicitly (e.g., **List**, **Instantiation**, and **Restatement**) while some others almost always appear with their markers (e.g., subtypes of **Condition**). Distributions of discourse cues

³We made a 2-d matrix with connective type vs. relation type dimensions and added 1 to the frequency appearing in each cell. Then we calculated $\frac{p(\text{cue}|r)}{p(\text{cue})}$ according to the resulting frequency table.

also differ to a similar extent between implicit and explicit occurrences, as relation senses and the discourse cues are highly correlated.

A smaller set of connectives appears to have been employed by annotators for the implicit relations. There are two possible reasons for this: first, some connectives such as “if” are markers of relations which cannot easily be expressed without a discourse connective. (For example, “if” is only used for explicit conditionals, and conditional discourse relations are expressed almost always with an explicit connector, so that no implicit “if” was annotated). A second possible reason for a connective not to appear frequently in the implicit annotations is if there exists a connective which is a better cue, or is much more frequent and has a similar function. An interesting case is the connective “when” which appears only a few times implicitly. One type of relation that this connective marks is the **reason** relation, which is very frequent in both implicit and explicit instances. The strongest marker of the **reason** relation is the connective “because” (11.80 strength), which makes it a better candidate when annotating implicit **reason** relations, compared to “when” (1.13 strength).

3.2 The Most Reliable Cues

The first thing we wanted to investigate by looking at the table of strength measurements was to find out which of the 95 connective types under study could most reliably mark a particular relation sense. To do this, we first looked at the strength measurements for frequent connectives. Among the 20 most frequent connectives in the corpus, a few showed a high strength score: “for example” for the **Instantiation** relation (42.17), “although” and “though” for the **expectation** relation (23.34 and 18.44), and “so” for the **result** relation (20.36). The highly frequent connectives “and” and “but” are associated with relatively small specificity scores (distributed strength) over a number of relation senses.

We found that 45 out of 95 connective types are used most frequently in some relation which is not the most specific relation they mark. Table 1 shows the strength scores and frequencies of six such connectives. It suggests that a number of relation instances including these connectives are not strongly marked. For example, “while” is used in many places as the connective of a **Synchrony** relation, but the negative bias in its meaning makes it a more reliable cue for an **opposition** relation, and the **Synchrony** relation is most reliably marked with “when” and “as”. Nevertheless, there is a subset of fairly frequent connectives which are associated with a very high strength to mark specific relation types. It includes

Connective	Most frequent relation	Strongest marking
and	Conjunction (2724, 3.04)	List (211, 3.71)
but	Juxtaposition (640, 6.54)	Contra-expectation (497, 7.20)
however	Juxtaposition (90, 6.01)	Contra-expectation (71, 6.72)
indeed	Conjunction (55, 1.57)	Specification (33, 24.39)
nor	Conjunction (27, 1.61)	Conjunctive (5, 11.15)
while	Synchrony (242, 3.72)	Opposition (91, 5.16)

Table 1: Comparison between the most frequent relation that a connective marks and the relation it marks with the highest strength (numbers in the brackets are the frequency of use and the strength of the connective for that relation, respectively).

“instead” for the chosen alternative (71.44), “or” for conjunctive (63.31) and “unless” for disjunctive (61.96). These connectives can be distinguished as very strong discourse markers with respect to the PDTB hierarchy of senses.

3.3 The Strongly Marked Relations

In the next step, we looked at different relation types to see which are most reliably marked by the connectives. We found that 12 out of 30 relation senses are most frequently marked with some connective that was not the most specific marker of that relation. Table 2 shows statistics for a number of such relations. In some cases, the strength associated with the typical marker is not very different from the maximum strength value obtained over all connectives for that relation. For example, **Conjunction** relations are usually marked by “and”, which exhibits a fairly similar strength score to “also”, i.e., the strongest marker of the relation. Although usage of “also” is very specific to the **Conjunction** relation, the small $p(\text{cue}|r)$ results in a relatively weak link between the relation and the connective. For some relations there is a big difference between the strength of the most frequently used connective and that of the strongest connective. A good example is **contra-expectation**, which in most cases appears in the corpus with “but”, a very generally used connective with a distributed marking strength over a variety of relation types. This would suggest that this relation type is usually not very strongly marked (as it could be marked by the use of “still” for example). We also investigated the variance of the strength values obtained over connective types for a particular relation. Interestingly, we found that the smallest variance of strength values was again obtained by the **Conjunction** relation, the most frequent relation in the corpus for which a number of connectives are used. We could imagine that if the **Conjunction** relation was divided into two or several subtypes (one might get help from the instances in which “also” is used to see whether a more specific relation sense can be considered), then each of those subtypes would be associated to a significantly greater strength of their cues.

Relation	Most frequent connective		Strongest marker	
Contra-expectation	but	(497, 7.20)	still	(81, 12.22)
Opposition	but	(177, 5.04)	on the other hand	(10, 8.16)
Factual present	if	(77, 6.18)	if then	(10, 14.86)
Pragmatic concession	but	(9, 1.06)	nevertheless	(6, 16.80)
Pragmatic contrast	but	(31, 3.18)	insofar as	(1, 4.93)
Conjunction	and	(2724, 3.04)	also	(1736, 3.50)
List	and	(211, 3.71)	finally	(8, 6.94)
Synchrony	when	(595, 5.64)	as	(544, 6.59)

Table 2: Comparison between the most frequent connective that marks a relation and the strongest marker of it (numbers in the brackets are the frequency of use and the strength of the connective for that relation, respectively).

4 Discussion and Conclusions

We reported examples of our measurements of discourse connective strength in reflecting relational senses. In this section, we will discuss how looking at the strength of discourse markers could be helpful in studies about discourse relations.

4.1 Development of the Corpora

Recent research on discourse processing, like other linguistic studies, is paying considerable attention to the corpus analysis. For this, a number of multi-purpose corpora of discourse-level annotated data, such as PDTB (Prasad et al., 2008) and RST-DTB (Carlson et al., 2003), have been developed. There are many theoretical and technical issues that need to be considered in developing such databases, some of which we think are relevant to our study of discourse markers:

Relation senses are not easy to define, especially when a corpus is being developed for a variety of research interests. Since discourse markers are the most important features one can use for defining (or choosing among previously-defined) discourse relation senses, statistics such as the strength of the cue become important. For example, the strength values of discourse cues marking the **Conjunction** relation are rather low. A different or more fine-grained division into subtypes for this relation might be worth considering.

Cross-corpora checking of the taxonomies (used for labeling discourse relations) could be useful in order to refine relation sense hierarchies. Van Dijk (2004) suggests that a discourse relation could either be of a *functional* type to establish intentional coherence between propositions in a text (e.g., the one proposition is a generalization / explanation / specification of the other), or of a referential type which expresses some extensional coherence between facts underlying the sentences (e.g., the facts stand in some causal / conditional / temporal relationship). He adds that these two types of relations have been confused in the literature (e.g., in RST Corpora) and need to be distinguished from one another. We believe that marker strength is potentially good means of studying the fine-grained classification of discourse relations, to distinguish for example between intensional and extensional coherence. Comparison of the relations tagged in two corpora with respect to the cue strength measurements might be helpful to find the overlaps or variance between relation sense definitions.

Implicit vs. explicit annotations of discourse relations are so far done simply according to the presence of *any* discourse connective. We would expect that in the near future many theoretical studies about discourse comprehension will be carried out on the basis of the available annotations. In such studies, the implicitness of a particular relation in a text should not be investigated solely in terms of the presence of a discourse marker. The markedness would strongly rely on the strength of the link between the relation type and the applied discourse cue and should be treated as a continuous feature. For example, **Reason** relations in the corpus which include “and” as their connective, are not really explicit causal relations, rather the causality is left implicit in the content of the arguments (this could further inform recent studies such as the one by Asr and Demberg (2012)).

4.2 Automatic Identification of Discourse Relations

Another aspect which is particularly important for computational linguists and NLP researchers is to develop a methods for automatically identifying discourse relations in a given text or utterance – which happens after defining a set of desired relation senses. We would suggest consideration of the following points both for human annotators and for development of automatic tools:

Discourse cues should be looked at with respect to their specificity, e.g., the measure we proposed to determine the marking strength of a word group. Every phrase or word in a

discourse relation could be counted as a cue, especially, when typically closed-class discourse connectives are not present. One example class of such markers are implicit causality verbs whose presence in a sentence can mark an upcoming reason (Asr and Demberg, 2012; Rohde and Horton, 2010). Another example is the presence of negation and / or downward-entailing verbs as a cue for an upcoming **chosen alternative** relation (Webber, 2012). Further examples include AltLex (Prasad et al., 2010), namely, alternative lexicalization of specific relations (e.g., “the reason is that...”) which might even be stronger markers than many of conjunctions such as “and” or “but”. The strength measure we proposed in this paper can be applied to any of these classes of cues, regardless of the syntactic differences. Only a strong cue can trigger expectation for a semantic/pragmatic relation between statements, thus, coherence of a text. On the other hand, mere presence of a sentence connective is a matter of textual cohesion (Halliday and Hasan, 1976).

Features for identification of relations can range from blind and coarse-grain properties of the propositional arguments (e.g, temporal focus of the events) to very fine-grained properties of the included discourse cues. We showed that the strength of the cue is a meaningful term in a simple probabilistic modeling of relation identification. Strength values can be calculated directly from the distribution of the discourse cues in a given corpus. Indeed, such a term should be used in a clear formulation along with the prior probability of the relation, i.e., the general expectation for a particular relation. Researchers have examined the classification of explicit and implicit discourse relations by only looking at the most typical relation that a discourse connective marks and obtained good accuracy for a coarse classification (Pitler et al., 2008; Zhou et al., 2010). To get an acceptable result for identification of fine-grained relation senses, one should definitely look into the strength of the discourse cue as some of them might not be reliable markers in a given context. Furthermore, it has been found that implicit relations are very difficult to classify correctly when learning only on explicit discourse relations (Sporleder, 2008). We would expect that weakly marked relations are similar to the unmarked relations; hence, one could possibly make use of this subset of explicit relations as training data for identification of implicit relations and get a different result.

4.3 Conclusions

This paper suggests a measure for the strength of a discourse cue in terms of its association with a specific discourse relation. We calculate this cue strength for discourse connectors and discourse relations as annotated in the Penn Discourse Treebank. We propose that such measurements are needed to understand how explicitly a discourse relation is marked in a text and what types of relations can be identified reliably by the use of their specific markers. Our findings also encourage the usage of a measure of cue strength in order to refine and develop robust annotations of discourse relation senses. We believe that theoretical as well as application-based studies in the field should in one way or another look into the strength of the link between the specific usage of words and phrases in a text and the type of coherence relation they reflect. Our preliminary findings can count as a trigger for future studies on discourse relations, the formalism and automated methods to identify them with respect to different types of discourse markers.

References

- Asr, F. T. and Demberg, V. (2012). Implicitness of discourse relations. In *Proceedings of COLING*, Mumbai, India.
- Carlson, L., Marcu, D., and Okurowski, M. (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. *Current and new directions in discourse and dialogue*, pages 85–112.
- Halliday, M. and Hasan, R. (1976). *Cohesion in English*. Longman (London).
- Müller, S. (2005). Discourse markers in native and non-native english discourse. *Pragmatics and Beyond*, 138:1–297.
- Pitler, E., Raghupathy, M., Mehta, H., Nenkova, A., Lee, A., and Joshi, A. (2008). Easily identifiable discourse relations. *Technical Reports (CIS)*, page 884.
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., and Webber, B. (2008). The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, pages 2961–2968.
- Prasad, R., Joshi, A., and Webber, B. (2010). Realization of discourse relations by other means: alternative lexicalizations. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 1023–1031.
- Rohde, H. and Horton, W. (2010). Why or what next? eye movements reveal expectations about discourse direction. In *Proceedings of 23rd Annual CUNY Conference on Human Sentence Processing*, pages 18–20.
- Sanders, T., Spooren, W., and Noordman, L. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, 15(1):1–35.
- Sporleder, C. (2008). Lexical models to identify unmarked discourse relations: Does WordNet help? *Lexical-Semantic Resources in Automated Discourse Analysis*, page 20.
- Stede, M. (2011). Discourse processing. *Synthesis Lectures on Human Language Technologies*, 4(3):1–165.
- Van Dijk, T. (2004). From text grammar to critical discourse analysis. *A brief academic autobiography. version, 2*.
- Webber, B. (2012). Alternatives and extra-propositional meaning. PASCAL 2 Invited Talk at ExProM Workshop, Jeju Island, Korea.
- Zhou, Z., Xu, Y., Niu, Z., Lan, M., Su, J., and Tan, C. (2010). Predicting discourse connectives for implicit discourse relation recognition. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 1507–1514.