

Using Syntax to Disambiguate Explicit Discourse Connectives in Text*

Emily Pitler and Ani Nenkova
Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
epitler, nenkova@seas.upenn.edu

Abstract

Discourse connectives are words or phrases such as *once*, *since*, and *on the contrary* that explicitly signal the presence of a discourse relation. There are two types of ambiguity that need to be resolved during discourse processing. First, a word can be ambiguous between discourse or non-discourse usage. For example, *once* can be either a temporal discourse connective or a simply a word meaning “formerly”. Secondly, some connectives are ambiguous in terms of the relation they mark. For example *since* can serve as either a temporal or causal connective. We demonstrate that syntactic features improve performance in both disambiguation tasks. We report state-of-the-art results for identifying discourse vs. non-discourse usage and human-level performance on sense disambiguation.

1 Introduction

Discourse connectives are often used to explicitly mark the presence of a discourse relation between two textual units. Some connectives are largely unambiguous, such as *although* and *additionally*, which are almost always used as discourse connectives and the relations they signal are unambiguously identified as comparison and expansion, respectively. However, not all words and phrases that can serve as discourse connectives have these desirable properties.

Some linguistic expressions are ambiguous between DISCOURSE AND NON-DISCOURSE USAGE. Consider for example the following sentences containing *and* and *once*.

- (1a) Selling picked up as previous buyers bailed out of their positions *and* aggressive short sellers—anticipating further declines—moved in.
- (1b) My favorite colors are blue *and* green.
- (2a) The asbestos fiber, crocidolite, is unusually resilient *once* it enters the lungs, with even brief exposures to it causing symptoms that show up decades later, researchers said.
- (2b) A form of asbestos *once* used to make Kent cigarette filters has caused a high percentage of cancer deaths among a group of workers exposed to it more than 30 years ago, researchers reported.

In sentence (1a), *and* is a discourse connective between the two clauses linked by an elaboration/expansion relation; in sentence (1b), the occurrence of *and* is non-discourse. Similarly in sentence (2a), *once* is a discourse connective marking the temporal relation between the clauses “The asbestos fiber, crocidolite is unusually resilient” and “it enters the lungs”. In contrast, in sentence (2b), *once* occurs with a non-discourse sense, meaning “formerly” and modifying “used”.

The only comprehensive study of discourse vs. non-discourse usage in written text¹ was done in the context of developing a complete discourse parser for unrestricted text using surface features (Marcu, 2000). Based on the findings from a corpus study, Marcu’s parser “ignored both cue phrases that had a sentential role in a majority of the instances in the corpus and those that were too ambiguous to be explored in the context of a surface-based approach”.

The other ambiguity that arises during discourse processing involves DISCOURSE RELATION SENSE. The discourse connective *since* for

¹The discourse vs. non-discourse usage ambiguity is even more problematic in spoken dialogues because there the number of potential discourse markers is greater than that in written text, including common words such as *now*, *well* and *okay*. Prosodic and acoustic features are the most powerful indicators of discourse vs. non-discourse usage in that genre (Hirschberg and Litman, 1993; Gravano et al., 2007)

This work was partially supported by NSF grants IIS-0803159, IIS-0705671 and IGERT 0504487.

instance can signal either a temporal or a causal relation as shown in the following examples from Miltsakaki et al. (2005):

- (3a) There have been more than 100 mergers and acquisitions within the European paper industry *since* the most recent wave of friendly takeovers was completed in the U.S. in 1986.
- (3b) It was a far safer deal for lenders *since* NWA had a healthier cash flow and more collateral on hand.

Most prior work on relation sense identification reports results obtained on data consisting of both explicit and implicit relations (Wellner et al., 2006; Soricut and Marcu, 2003). Implicit relations are those inferred by the reader in the absence of a discourse connective and so are hard to identify automatically. Explicit relations are much easier (Pitler et al., 2008).

In this paper, we explore the predictive power of syntactic features for both the discourse vs. non-discourse usage (Section 3) and discourse relation sense (Section 4) prediction tasks for explicit connectives in written text. **For both tasks we report high classification accuracies close to 95%.**

2 Corpus and features

2.1 Penn Discourse Treebank

In our work we use the Penn Discourse Treebank (PDTB) (Prasad et al., 2008), the largest public resource containing discourse annotations. The corpus contains annotations of 18,459 instances of 100 explicit discourse connectives. Each discourse connective is assigned a sense from a three-level hierarchy of senses. **In our experiments we consider only the top level categories:** Expansion (one clause is elaborating information in the other), Comparison (information in the two clauses is compared or contrasted), Contingency (one clause expresses the cause of the other), and Temporal (information in two clauses are related because of their timing). **These top-level discourse relation senses are general enough to be annotated with high inter-annotator agreement and are common to most theories of discourse.**

2.2 Syntactic features

Syntactic features have been extensively used for tasks such as argument identification: dividing sentences into elementary discourse units among which discourse relations hold (Soricut and Marcu, 2003; Wellner and Pustejovsky, 2007; Fisher and Roark, 2007; Elwell and Baldridge,

2008). Syntax has not been used for discourse vs. non-discourse disambiguation, **but it is clear from the examples above that discourse connectives appear in specific syntactic contexts.**

The syntactic features we used were extracted from the gold standard Penn Treebank (Marcus et al., 1994) parses of the PDTB articles:

Self Category **The highest node in the tree which dominates the words in the connective but nothing else.** For single word connectives, this might correspond to the POS tag of the word, however for multi-word connectives it will not. For example, the cue phrase *in addition* is parsed as (PP (IN In) (NP (NN addition))). While the POS tags of “in” and “addition” are preposition and noun, respectively, together the Self Category of the phrase is prepositional phrase.

Parent Category **The category of the immediate parent of the Self Category.** This feature is especially helpful for disambiguating cases similar to example (1b) above in which the parent of *and* would be an NP (the noun phrase “blue and green”), which will rarely be the case when *and* has a discourse function.

Left Sibling Category The syntactic category of the sibling immediately to the left of the Self Category. **If the left sibling does not exist, this feature takes the value “NONE”.** Note that having no left sibling implies that the connective is the first substring inside its Parent Category. In example (1a), this feature would be “NONE”, while in example (1b), the left sibling of *and* is “NP”.

Right Sibling Category The syntactic category of the sibling immediately to the right of the Self Category. **English is a right-branching language, and so dependents tend to occur after their heads. Thus, the right sibling is particularly important as it is often the dependent of the potential discourse connective under investigation.** If the connective string has a discourse function, then this dependent will often be a clause (SBAR). For example, the discourse usage in “*After* I went to the store, I went home” can be distinguished from the non-discourse usage in “*After* May, I will go on vacation” based on the categories of their right siblings.

Just knowing the syntactic category of the right sibling is sometimes not enough; experiments on the development set showed improvements by including more features about the right sibling.

Consider the example below:

- (4) NASA won’t attempt a rescue; instead, it will try to predict whether any of the rubble will smash to the ground

and where.

The syntactic category of “where” is SBAR, so the set of features above could not distinguish the single word “where” from a full embedded clause like “I went to the store”. In order to address this deficiency, we include two additional features about the contents of the right sibling, **Right Sibling Contains a VP** and **Right Sibling Contains a Trace**.

3 Discourse vs. non-discourse usage

Of the 100 connectives annotated in the PDTB, only 11 appear as a discourse connective more than 90% of the time: *although*, *in turn*, *afterward*, *consequently*, *additionally*, *alternatively*, *whereas*, *on the contrary*, *if and when*, *lest*, and *on the one hand...on the other hand*. There is quite a range among the most frequent connectives: *although* appears as a discourse connective 91.4% of the time, while *or* only serves a discourse function 2.8% of the times it appears.

For training and testing, we used explicit discourse connectives annotated in the PDTB as positive examples and occurrences of the same strings in the PDTB texts that were not annotated as explicit connectives as negative examples.

Sections 0 and 1 of the PDTB were used for development of the features described in the previous section. Here we report results using a **maximum entropy classifier**² using ten-fold cross-validation over sections 2-22.

The results are shown in Table 3. Using the string of the connective as the only feature sets a reasonably high baseline, with an f-score of 75.33% and an accuracy of 85.86%. Interestingly, using only the syntactic features, ignoring the identity of the connective, is even better, resulting in an f-score of 88.19% and accuracy of 92.25%. Using both the connective and syntactic features is better than either individually, with an f-score of 92.28% and accuracy of 95.04%.

We also experimented with combinations of features. It is possible that different connectives have different syntactic contexts for discourse usage. Including **pair-wise interaction features between the connective and each syntactic feature** (features like *connective=also-RightSibling=SBAR*) raised the f-score about 1.5%, to 93.63%. **Adding interaction terms between pairs of syntactic features** raises the f-score

²<http://mallet.cs.umass.edu>

| Features | Accuracy | f-score |
|----------------------------------|--------------|--------------|
| (1) Connective Only | 85.86 | 75.33 |
| (2) Syntax Only | 92.25 | 88.19 |
| (3) Connective+Syntax | 95.04 | 92.28 |
| (3)+Conn-Syn Interaction | 95.99 | 93.63 |
| (3)+Conn-Syn+Syn-Syn Interaction | 96.26 | 94.19 |

Table 1: Discourse versus Non-discourse Usage

slightly more, to 94.19%. These results amount to a 10% absolute improvement over those obtained by Marcu (2000) in his corpus-based approach which achieves an f-score of 84.9%³ for identifying discourse connectives in text. While bearing in mind that the evaluations were done on different corpora and so are not directly comparable, as well as that our results would likely drop slightly if an automatic parser was used instead of the gold-standard parses, syntactic features prove highly beneficial for discourse vs. non-discourse usage prediction, as expected.

4 Sense classification

While most connectives almost always occur with just one of the senses (for example, *because* is almost always a Contingency), a few are quite ambiguous. For example *since* is often a Temporal relation, but also often indicates Contingency.

After developing syntactic features for the discourse versus non-discourse usage task, we investigated whether these same features would be useful for sense disambiguation.

Experiments and results We do classification between the four senses for each explicit relation and report results on ten-fold cross-validation over sections 2-22 of the PDTB using a Naive Bayes classifier⁴.

Annotators were allowed to provide two senses for a given connective; in these cases, we consider either sense to be correct⁵. Contingency and Temporal are the senses most often annotated together. The connectives most often doubly annotated in the PDTB are *when* (205/989), *and* (183/2999), and *as* (180/743).

Results are shown in Table 4. The sense classification accuracy using just the connective is already quite high, 93.67%. Incorporating the syntactic features raises performance to 94.15% accu-

³From the reported precision of 89.5% and recall of 80.8%

⁴We also ran a MaxEnt classifier and achieved quite similar but slightly lower results.

⁵Counting only the first sense as correct leads to about 1% lower accuracy.

| Features | Accuracy |
|--|--------------|
| Connective Only | 93.67 |
| Connective+Syntax+Conn-Syn | 94.15 |
| Interannotator agreement on sense class (Prasad et al., 2008) | 94 |

Table 2: Four-way sense classification of explicit

racy. While the improvement is not huge, note that we seem to be approaching a performance ceiling. The human inter-annotator agreement on the top level sense class was also 94%, suggesting further improvements may not be possible. We provide some examples to give a sense of the type of errors that still occur.

Error Analysis While Temporal relations are the least frequent of the four senses, making up only 19% of the explicit relations, **more than half of the errors involve the Temporal class**. By far the most commonly confused pairing was **Contingency relations being classified as Temporal relations, making up 29% of our errors**.

A random example of each of the most common types of errors is given below.

- (5) *Builders get away with using sand and financiers junk [when] society decides it's okay, necessary even, to look the other way.* Predicted: Temporal Correct: Contingency
- (6) *You get a rain at the wrong time [and] the crop is ruined.* Predicted: Expansion Correct: Contingency
- (7) *In the nine months, imports rose 20% to 155.039 trillion lire [and] exports grew 18% to 140.106 trillion lire.* Predicted: Expansion Correct: Comparison
- (8) *[The biotechnology concern said] Spanish authorities must still clear the price for the treatment [but] that it expects to receive such approval by year end.* Predicted: Comparison Correct: Expansion

Examples (6) and (7) show the relatively rare scenario when *and* does not signal expansion, and Example (8) shows *but* indicating a sense besides comparison. In these cases where the connective itself is not helpful in classifying the sense of the relation, it may be useful to incorporate features that were developed for classifying implicit relations (Sporleder and Lascarides, 2008).

5 Conclusion

We have shown that using a few syntactic features leads to state-of-the-art accuracy for discourse vs. non-discourse usage classification. Including syntactic features also helps sense class identification, and we have already attained results at the level of human annotator agreement. These results taken

together show that explicit discourse connectives can be identified automatically with high accuracy.

References

- R. Elwell and J. Baldridge. 2008. Discourse connective argument identification with connective specific rankers. In *Proceedings of the International Conference on Semantic Computing, Santa Clara, CA*.
- S. Fisher and B. Roark. 2007. The utility of parse-derived features for automatic discourse segmentation. In *Proceedings of ACL*, pages 488–495.
- A. Gravano, S. Benus, H. Chavez, J. Hirschberg, and L. Wilcox. 2007. On the role of context and prosody in the interpretation of 'okay'. In *Proceedings of ACL*, pages 800–807.
- J. Hirschberg and D. Litman. 1993. Empirical studies on the disambiguation of cue phrases. *Computational linguistics*, 19(3):501–530.
- D. Marcu. 2000. The rhetorical parsing of unrestricted texts: A surface-based approach. *Computational Linguistics*, 26(3):395–448.
- M.P. Marcus, B. Santorini, and M.A. Marcinkiewicz. 1994. Building a large annotated corpus of english: The penn treebank. *Computational Linguistics*, 19(2):313–330.
- E. Miltsakaki, N. Dinesh, R. Prasad, A. Joshi, and B. Webber. 2005. Experiments on sense annotation and sense disambiguation of discourse connectives. In *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT 2005)*.
- E. Pitler, M. Raghupathy, H. Mehta, A. Nenkova, A. Lee, and A. Joshi. 2008. Easily identifiable discourse relations. In *COLING, short paper*.
- R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. 2008. The penn discourse treebank 2.0. In *Proceedings of LREC'08*.
- R. Soricut and D. Marcu. 2003. Sentence level discourse parsing using syntactic and lexical information. In *HLT-NAACL*.
- C. Sporleder and A. Lascarides. 2008. Using automatically labelled examples to classify rhetorical relations: An assessment. *Natural Language Engineering*, 14:369–416.
- B. Wellner and J. Pustejovsky. 2007. Automatically identifying the arguments of discourse connectives. In *Proceedings of EMNLP-CoNLL*, pages 92–101.
- B. Wellner, J. Pustejovsky, C. Havasi, A. Rumshisky, and R. Sauri. 2006. Classification of discourse coherence relations: An exploratory study using multiple knowledge sources. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*.