Uniform Information Density at the Level of Discourse Relations: Negation Markers and Discourse Connective Omission

Fatemeh Torabi Asr & Vera Demberg
Saarland University
`fatemeh|vera@coli.uni-saarland.de`

Abstract

About half of the discourse relations annotated in Penn Discourse Treebank (Prasad et al., 2008) are not explicitly marked using a discourse connective. But we do not have extensive theories of when or why a discourse relation is marked explicitly or when the connective is omitted. Asr and Demberg (2012a) have suggested an information-theoretic perspective according to which discourse connectives are more likely to be omitted when they are marking a relation that is expected or predictable. This account is based on the Uniform Information Density theory (Levy and Jaeger, 2007), which suggests that speakers choose among alternative formulations that are allowed in their language the ones that achieve a roughly uniform rate of information transmission. Optional discourse markers should thus be omitted if they would lead to a trough in information density, and be inserted in order to avoid peaks in information density. We here test this hypothesis by observing how far a specific cue, negation in any form, affects the discourse relations that can be predicted to hold in a text, and how the presence of this cue in turn affects the use of explicit discourse connectives.

1 Introduction

Discourse connectives are known as optional linguistic elements to construct relations between clausal units in text: in the Penn Discourse Treebank (PDTB Prasad et al. 2008), only about half of the annotated discourse relations are marked in the text by a discourse connective. In the remaining cases, the discourse relation can still be recovered without an explicit marker. Consider for example the sentences below, which stand in a causal relationship:

(1) a. John did not go to the concert. He was ill.
     b. John did not go to the concert, because he was ill.

The question we would like to discuss in this paper is whether there is a principled explanation of when such optional discourse markers are inserted by speakers / writers, and when they are omitted.

In Grician pragmatics the *maxim of quantity* holds that speakers should make their contribution as informative as is required for the listeners to grasp the message, but not more informative than is required (Grice, 1975). The *Uniform Information Density* theory (UID, Levy and Jaeger 2007) further refines this notion with respect to how information can be transferred optimally from a speaker to a hearer: in order to optimally use the comprehenders channel capacity, the speaker should distribute the information uniformly across the utterance, in a manner that approximates channel capacity. In particular, among alternative linguistic formulations of the same content, the one formulation that conveys the information more uniformly should be preferred.

The amount of information conveyed by a word is quantified in terms of its *Surprisal* (Hale, 2003). Surprisal of a word w_i in a sentence is calculated based on its conditional probability given the preceding context:

$$s(w_i) = -\log p(w_i|w_{1...i-1}) \qquad (1)$$

From the perspective of the speaker, the Uniform Information Density hypothesis predicts that the amount of surprisal should be held roughly equal from word to word. If information is transmitted at a rate close to channel capacity, information transmission is optimal because the maximal amount of information can be transmitted while avoiding comprehension problems.

Evidence that speakers indeed behave in this way, and choose among meaning-equivalent alternatives the ones that correspond to a more uniform rate of information transmission has been provided by a range of experimental and corpus-based studies at the level of spoken word duration and articulation (Aylett and Turk, 2004; Buz et al., 2014), morphology (Kurumada and Jaeger, 2013), syntax (Jaeger, 2010), lexical choices (Piantadosi et al., 2011; Mahowald et al., 2013), referring expressions (Tily and Piantadosi, 2009; Kravtchenko, 2014), and across levels, e.g., effect from syntax on spoken word durations (Demberg et al., 2012). We here investigate whether UID can also explain discourse-level phenomena, such as the insertion vs. omission of discourse connectives.

Some first evidence for this hypothesis comes from Asr and Demberg (2012a), who looked into the discourse-relation annotated Penn Discourse Treebank (PDTB, Prasad et al. 2008) corpus to explore **what relation senses** tend to be expressed without discourse connectives. Asr and Demberg observe that discourse relations which are predictable given general cognitive biases in relation interpretation, such as continuity and causality are more likely to be expressed without an explicit discourse connective, while unpredictable (discontinuous, adversative or temporally backward) relations tend to be marked with an explicit connective.

These observations are in line with the general predictions of theories about efficient language production such as Grice's maxim of quantity and the UID account. However, the UID more specifically suggests that even markers of generally unexpected relations may be subject to omission if there are other strong cues in the local context, which make the relation predictable. The present paper addresses this gap by specifically looking into how far a generally unexpected discourse relation, `chosen alernative`, can be expressed without its typical discourse connective *instead* when a good cue is present in the first argument of the discourse relation (the "good cue" here will be negation).

After providing some background about the Penn Discourse Treebank and discourse connectives in section 2, we propose how to calculate surprisal for discourse relations in section 3. Our experiments in section 4 first focus on the relational information encoded in negation. We investigate what types of relations benefit from negation as a statistically licensed marker. Second, under the notion of UID, we predict that the optional marker of a discourse relation (*instead* for the `Chosen alternative` relation) should be dropped as a function of linguistic features in the first argument (here, negative polarity), which are predictive of the relation sense. We find not only that the polarity of a sentence changes the distribution of the discourse relations that it makes with the context, but also that the relational surprisal calculated based on this feature is a predictor of the connective ellipsis.

2   Background and related work

This section introduces discourse relations, specifically according to the definition employed in annotation of the PDTB. We also sketch an overview of the studies on the linguistic markers of discourse relations such as sentence connectives and clausal features. Finally, we focus on the studies that look at the question of implicitness.

2.1   Discourse relations in PDTB

PDTB is the largest resource for discourse related studies published so far. It contains about fifty thousand annotated relations on text from Wall Street Journals. Relations are considered for pairs of clauses connected by a discourse connective, as well as between neighboring sentences which are not connected by any discourse cue. If two clauses are joined by a discourse connective the boundaries of the arguments are annotated and a label indicating the relation sense is assigned. Otherwise, the annotators were asked to first see whether any discourse connective could artificially be inserted between the two arguments
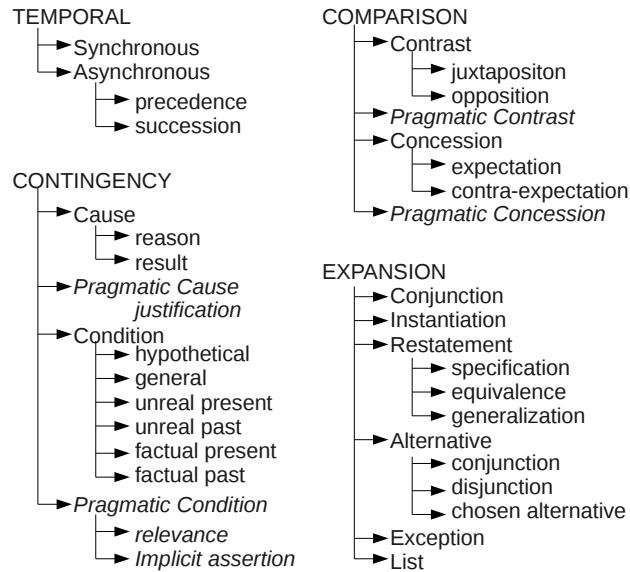
TEMPORAL
→ Synchronous
→ Asynchronous
→ precedence
→ succession

CONTINGENCY
→ Cause
→ reason
→ result
→ *Pragmatic Cause*
  justification
→ Condition
→ hypothetical
→ general
→ unreal present
→ unreal past
→ factual present
→ factual past
→ *Pragmatic Condition*
→ *relevance*
→ *Implicit assertion*

COMPARISON
→ Contrast
→ juxtapositon
→ opposition
→ *Pragmatic Contrast*
→ Concession
→ expectation
→ contra-expectation
→ *Pragmatic Concession*

EXPANSION
→ Conjunction
→ Instantiation
→ Restatement
→ specification
→ equivalence
→ generalization
→ Alternative
→ conjunction
→ disjunction
→ chosen alternative
→ Exception
→ List

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

and then apply the same procedure to annotate a relation sense accordingly. In the latter case the relation is called *implicit* given that the discourse connective is absent from the original text [1]. Annotation of the relation senses is according to a hierarchy of coarse to fine-grained semantic categories depicted in Figure 1. For example, a `chosen alternative` relation is a very specific sense located in depth three of the hierarchy, and it is used when the connective indicates that its two arguments denote alternative situations but only one (Arg2) has occurred.

(2)   *[No price for the new shares has been set.] Instead, [the companies will leave it up to the marketplace to decide.]* — `EXPANSION.Alternative.chosen alternative`

Some relations are annotated with less specificity due to the disagreement between annotators, or with two different sense labels when both relations are conveyed simultaneously [2].

2.2   Discourse connectives

Discourse connectives are words or expressions which connect neighboring sentences and clauses in a text. Starting by the work of Halliday and Hasan (1976) these linguistic elements have been identified as cohesion devices. Later researchers argue that in addition to the surface connectivity, discourse connectives provide constraints on how readers should relate the content of the linked sentences (Trabasso et al., 1982; Blakemore, 1992; Sanders and Noordman, 2000). Experimental studies have proven the effect of these cues on various aspect of sentence processing (Caron et al., 1988; Millis and Just, 1994; Murray, 1995; Köhne and Demberg, 2013). For example, Murray (1995) showed that each category of connectives, such as causal vs. adversative, triggers a different expectation of the text continuation and if a relation is expressed with an infelicitous connective, readers face comprehension difficulty. Köhne and Demberg (2013) carried out an online reading study that revealed such effects show up very early during reading the second argument of a relation and confirmed that the online influence of a sentence connective can be as fine grained as triggering lexical predictions.

While discourse connectives are know as the best markers of discourse relations, empirical studies such as Knott (1996) and Asr and Demberg (2012b; 2013) indicate that connective types vary a lot in

---

[1] Other than explicit and implicit discourse relations, PDTB contains a set of other types of relations which are not considered here.

[2] Interested readers are referred to Prasad et al. (2008) for more details on the annotation guidelines.

terms of the granularity and amount of relational information they encode and this goes beyond the typical coarse-grained categorization of connectives. In particular, Asr and Demberg (2013) show that *instead* and *because* are highly informative (or strongly constraining) markers which should play an important role in identification of very specific discourse relations, whereas *and* and *but* occur in a wide range of contexts, thus do not disambiguate the relation down to a very specific sense.

2.3   Other relational cues

Even when connectives are absent or do not exhibit enough information about the relation sense between two consecutive discourse units, readers can still infer relations by relying on their world-knowledge and the information encoded in the arguments of the relation; See an implicit *reason* relation in the following example.

(3)   John admires Mary. She plays the piano very well.

Rohde and Horton (2010, 2014) provide more evidence for the incrementality of discourse relation processing within a novel visual world experiment, which specifically shows that people predict the relation between two sentences as soon as they encounter a cue, be it the discourse connective or other linguistic cues yet within the first clause. Implicit causality verbs, such as *admire* in the above example, trigger expectation for a causal continuation and given the time course of the effect in Rohde and Horton's experiment one can infer that the connective (in this case *because*) would then be a redundant operator. Recently, a few systematic corpus studies have been conducted on the discovery of other types of relational markers in natural text (Prasad et al., 2010; Das and Taboada, 2013; Duque, 2013; Webber, 2013). The obtained annotations throughout these studies indicate that, in fact, a lot of discourse relations benefit from other types of cues besides explicit discourse connectives. Furthermore, machine learning attempts for detection of implicit discourse relations reveal that lexical, syntactic and clause-level properties of the arguments can determine the relation sense with good accuracy, at least for a coarse-grained classification (Pitler et al., 2009; Lin et al., 2009; Zhou et al., 2010; Park and Cardie, 2012; Rutherford and Xue, 2014).

2.4   Implicitness

Regarding the question of implicitness, Asr and Demberg (2012a, 2013) study causal and continuous discourse relations which based on cognitive theories and lab experiments tend to be eagerly inferred by readers when sentences are encountered consecutively. In both studies, Asr and Demberg extract different senses of relations from PDTB and find a higher degree of implicitness (proportion of the implicit to the explicit occurrences) for causal and continuous relation senses compared with other types of relations, e.g. comparison or backward temporal relations. This finding has been interpreted as an evidence that writers consider the reader side default inferential biases during language production. Asr and Demberg (2012a) also investigate the effect of implicit causality verbs as a linguistic cue in Arg1 of the relation. Regarding the observations of Rohde and Horton (2010) about the comprehension of causal relations, they predict a higher degree of implicitness for `CONTINGENCY.Cause.reason` relations containing implicit causality verbs. Analysis of these relations in PDTB fails to show any significant correlation between the two factors which is attributed to the noise involved with the automatic extraction and sense disambiguation of the targeted verbs. Beside that, the verb type can simply be a non-salient feature at the discourse level and in particular implicit causality verbs might generate expectation for causal relations only in a particularly constrained context. In fact, the lab experiments regarding the effect of these verbs on comprehension have benefited from very specific form of text, i.e., a simple and short narration about two protagonists like in 2.3 which makes the stimuli very different from sentences in expository text. Furthermore, the majority of previous lab findings about this class of verbs support the predictability of a following referent rather than the relation sense.

Patterson and Kehler (2013) train a classifier on the linguistic features of the arguments of PDTB relations to distinguish between implicit and explicit occurrences. The results suggests that implicit and

explicit relations are indeed different, i.e., it is possible to predict based on the arguments whether a connective is required or not. Nevertheless, their findings lack interpretability: while the classifier can suggest with acceptable accuracy when a discourse connective is necessary (86.6% in binary classification), it does not provide any insight why this either case is preferred. They also run a comparison task on Amazon Mechanical Turk to see how human judgment differs from the automatic classification on the same test data set. They find that human preference in selecting between implicit/explicit variations of the relations aligns with the original text even less often (68% accuracy). The authors propose that it could be due to genre-specific editorial regulations that is applied in Wall Street Journal, e.g., use of sentence initial *but*, that is not recommended in prescriptive grammar books and, in turn, not preferred by Amturk subjects. From a psycholinguistic perspective, there is always a difference between decisions that human language users subconsciously make in writing and what they might prefer in an explicit judgment task, because a different cognitive mechanism is activated in each task.

2.5 Our focus: Sentence polarity

This paper presents an attempt to investigate the interaction of strong discourse relation cues (which are not considered as traditional discourse markers) with presence/absence of the sentence connectives. We look into the use of negation words in the first argument of a relation as an incrementally available cue for predicting the relation between a sentence and the upcoming discourse unit.

Webber's manual analysis of the `chosen alternative` relations in PDTB, as well as a corpus of *instead sentences*, reveals that the first argument of this type of relation usually contains some type of negation or a downward entailing verb (Webber, 2013). Based on this observation and the fact that negation (as long as the scope is not resolved) is a relatively easy feature to automatically detect in text, we decided to look into this feature as a case study of UID at the level of inter-sentential coherence relations. Another motivation for studying negation comes from the emphasized importance given to the relation polarity as a cognitively plausible dimension for classification of discourse relations (Sanders, 1997)[3]. Hence, in Section 5, we will examine:

- whether the presence of explicit negation words in a sentence changes the distribution of the discourse relations and in particular is statistically licensed as a feature of the `chosen alternative` relations.

- whether the presence/absence of the connectives can be explained according to the UID, i.e., predicted by relational surprisal which in turn is calculated based on the negation feature extracted from Arg1.

3 Surprisal at the discourse connective

The standard formulation of surprisal at a particular word w is its negative log probability in context: $-\log P(w|context)$. We can then calculate, for example, the syntactic surprisal of a word in context as the difference between the sum of the probabilities of all syntactic trees spanning words $w_1..w_{i-1}$ and the sum of the probabilities of all trees additionally including word w_i:

$$S_{syntactic}(w_i) = -\log \sum_{T \in Trees} P(T, w_1..w_i) - -\log \sum_{T \in Trees} P(T, w_1..w_{i-1})$$

(see Levy, 2008; Demberg and Keller, 2008). Similarly, we can define discourse relation surprisal as

$$S_{relational}(w_i) = -\log \sum_{R \in DiscRels} P(R, w_1..w_i) - -\log \sum_{R \in DiscRels} P(R, w_1..w_{i-1})$$

---

[3]Note, however, that negation in the surface is not necessarily equivalent to negative relational polarity. For example, "Mary loves John, but she pretends to ignore him." is a negative polarity relation without utilizing any covert negation, whereas, "Mary doesn't love John and she pretends to ignore him" is a positive polarity relation including some negation.

General Intuition. Discourse relation surprisal quantifies how much a word w_i changes the distribution of all possible instances of discourse relations R. In this conceptualisation, all words can potentially convey some information about the discourse relations in the text. Discourse connectives and adverbials like *but*, *since* or *because* are treated the same way as other cues in the text — while they are generally very informative cues about upcoming discourse relations, almost all of the discourse connectives are ambiguous, i.e. they can occur as a marker of several discourse relations, and are therefore best treated probabilistically, just like other cues. The perspective of all words being potential cues for discourse relations allows us to account for how words in the arguments of a discourse relation can affect discourse relation inference beyond the connective. Cues such as negation, event modals, coreference patterns, temporal phrases, verb tense and modality are also known to play a role for discourse relation inference. We hypothesise that humans have certain general expectations about upcoming discourse relations (i.e., a preference for causal and continuous relations), and that these expectations are updated by incoming words that contribute to shape the anticipation of a discourse relation. For example, a negation would convey the discourse-relation relevant information that a likely upcoming discourse relation could be a chosen alternative, a contrast or an explanation.

Conceptual Simplification for Avoiding Data Sparsity. In this paper, we focus on information conveyed by a cue in the first argument and its effect on the presence of a discourse connective like *instead*. The connective should be omitted if the information it conveys is highly predictable given previous context, or if its presence would otherwise make non-optional words in the second argument highly predictable and non-informative. The latter effect is much harder to estimate with the amount of data that we have; therefore, we will concentrate on the first part.

The information conveyed by a connective is thus:

$$S(connective) = -\log \sum_{R \in DiscRels} P(R, connective, arg1) - -\log \sum_{R \in DiscRels} P(R, arg1)$$

If the distribution of expected discourse relations does not change much by observing the connective, then surprisal at the discourse connective is very small, leading to a trough in information density. The connective should then be omitted. Hence, the interesting component in the above formula is $P(R|arg1)$, indicative of the information already available to predict the relation before encountering the discourse connective.

4 Experiment

In general, $P(R|arg1)$ can be computed for all possible values of relations R based on a set of features extracted from the first argument of the discourse relations in a reference corpus. In this study, we start off with a single feature within Arg1 of the relation. If as we predicted, connectives are used as modulator of relational surprisal, we should find for a given relation sense $R \in DiscRels$ that the connective $Conn$ marking R is present where $P(R|arg1)$ is small and that it should be omitted when $P(R|arg1)$ is high.

4.1 Negation in the arguments of discourse relations

A binary feature is defined indicating whether any of the following negation words is present in the Arg1 of a relation: {*not, n't, no, without, never, neither, none, non , nor, nobody, nothing*}. In PDTB, the Arg1 of an implicit relation is always the one that appears first in the text. In case of explicit relations, the order of arguments is reversed, when a sentence initial subordinating connective is used (that is, Arg2 is always the one directly following the connective in explicitly marked relations). Since our argument is based upon incremental processing of the relational cues, the analysis presented in this paper excludes relations with reversed arguments, 1920 instances in the corpus – we'd like to note however that same analysis leads to very similar results when reversed relations included. Among all implicit and explicit relations under analysis about 14% turn out to have some negation in their Arg1.
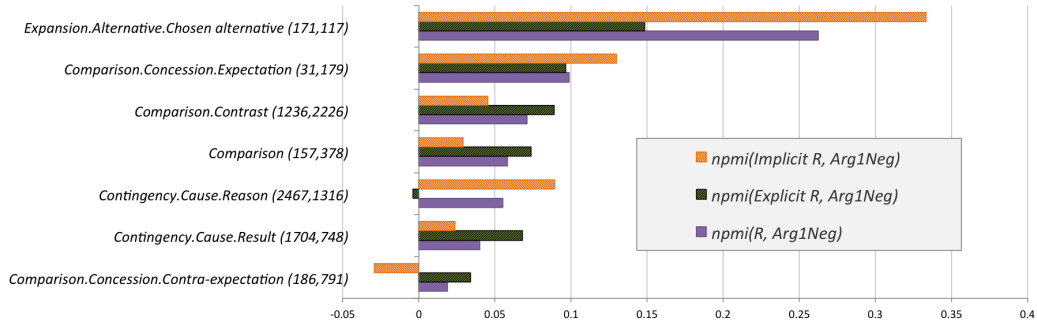
Figure 2: Relation senses showing positive $npmi$ with negation in Arg1 (all significant at $p < 0.001$ except the `contra-expectation`). Frequency of implicit and explicit occurrences of every relation sense is indicated in brackets.

To test the reliability of the automatic procedure in discovery of negative words, we compared our list of explicit `chosen alternative` relations with the list manually analyzed by Webber (2013) and discovered only 1 difference where our algorithm found a negation in the first argument but it was not considered by Webber as a marker of the relation. Webber also detected 5 influential negation words in the attribution of the relations, as well as, 5 negations in a larger context rather than in the Arg1 boundaries. We do not consider such cases for the matter of consistency, i.e., we focus on the linguistic cues inside Arg1. Finally, all relations are considered with their fine-grained senses annotated in the PDTB and if a relation is annotated with more than one sense in the corpus, we count it for every sense separately. We only look into the relations that have 30 or more implicit and 30 or more explicit instances in the corpus.

4.2 Correlation bw. polarity and relation senses

Regarding the investigation of the UID hypothesis, we first need to make sure that negation is statistically licensed as a feature of the `chosen alternative` relation. Merely knowing that such a cues occurs frequently with this type of relation is not enough, because negation might be a strong marker of some other relation(s) too. If other relation senses turn out to be more likely given the presence of negation, the connective in `chosen alternative` would not be subject to omission in order to keep a uniform discourse-level information density, or such a correlation would be expected to a lesser extent. If on the other hand, `chosen alternative` turns out to be the most likely relation sense given the negation cue, then it would make a perfect test case for investigation of the UID.

Similar to Asr and Demberg's analysis of discourse connectives strength, we calculate the normalized point-wise mutual information (Bouma, 2009) between the polarity of the first argument of a relation as a binary cue, and its sense:

$$npmi(R|C) = \frac{\log p(R)p(C)}{\log p(R,C)} - 1 \qquad (2)$$

This removes the unwanted effect of the raw frequency of the relations and provides us with a scaled measure to see what relations benefit from negation as a cue. Figure 2 shows the relation senses obtaining a positive $npmi$ with the *negation* cue in Arg1. Other relation senses either obtain a significantly negative score (e.g., `synchronous` which indicates that a negative polarity sentence in a text would least likely be followed by this relation) or a closed to zero score, i.e., no correlation. The `chosen alternative` relation, in particular, is located at the top, meaning that negation in Arg1 is highly predictive of this relation sense. Running the same analysis by considering only the implicit relations in the corpus reveals an even stronger pattern.

4.3 Relational surprisal as a predictor of implicitness

The posterior probability of the `chosen alternative` given the presence of a negation cue, $3.45 * 10^{-2}$, is much higher than its prior which is $8.44 * 10^{-3}$. Among the `chosen alternative` relations, absence of the discourse connective is positively correlated with the presence of negation: the likelihood of the connective being dropped is 75.6% for relations with some negation in Arg1, whereas, it is only 39% for the rest of `chosen alternative` relations ($p < 0.001$). This observation is consistent with our hypothesis: if a cue like negation is helpful to human comprehenders (or affects what speakers do) in order to infer a discourse relation, then the relation should be marked less often in presence of that cue.

Figure 2 indicates that two other relation senses, i.e., `expectation` and `reason` show similar patterns (though the UID effect is only significant for `reason`). On the other hand, `COMPARISON` and `COMPARISON.Contrast` relations show an opposite trend, i.e., while negation in Arg1 increases the likelihood of these relations, the connectives marking these relations tend to be dropped in the presence rather than absence of the negation feature (not a significant difference though). This raises a question: why should the presence of negation in Arg1 of `chosen alternative` correlate with its implicitness but not for some other relation types? It is quite possible that the necessity of the connective is affected by other factors not included in our analysis. An alternative explanation would be that negation is not a similarly salient feature for `contrast` relations as it is for `chosen alternative` relations. In fact, the `Contrast` relation is much more common, and negation does not affect its distribution (prior $1.01 * 10^{-1}$, posterior $1.34 * 10^{-1}$) as much as it affects the distribution of the `chosen alternative` relation. Finally, as pointed out before, discourse connectives differ a lot in terms of their relational information content, hence, presence/absence of the highly informative connectives such as *instead* and *because* is more of a UID related question compared with relatively less informative, i.e., ambiguous connectives such as *but* which is typically used in `Contrast` relations.

5   Discussion

Our analysis here was based on the predictability of a discourse relation given a local cue in the first argument of a discourse relation. Considering only a single feature in the first argument is a rather poor measure of estimating the real predictive effect of the words in the first argument of a discourse relation on the identity of the discourse relation, in particular in the absence of more sophisticated methods for determining negation scope in our approach. This method on average hence seriously under-estimates the predictability of discourse relations. Nevertheless, we found support for the hypothesis that negation is a good cue for the `chosen alternative` relation (Webber, 2013), and were furthermore able to show that the distribution of the discourse connectors marking this relation is consistent with the predictions of the Uniform Information Density hypothesis.

A second point to take into account when thinking about these results is that the insertion or omission of a discourse connector should not only depend on the amount of new information it conveys with respect to earlier cues in the text, but rather that its presence might also be determined by whether it is redundant with respect to information contained in the second argument (which may not be left out due to reasons of grammaticality): the prediction is that the optional linguistic element (the connector) should be omitted if that leads to a more uniform distribution of information across the utterance. It would be particularly interesting to connect this observation with the speaker's planning capacities: such considerations should be more difficult to take into account during language production, when the distance between two words that are partially redundant with respect to one another is larger.

6   Conclusions

Following the Uniform Information Density hypothesis, we predicted that optional markers of discourse relations should be used in cases when they convey a substantial amount of new information, and should be left out if the information they convey is predictable given other cues in the incrementally available

context. In order to investigate this hypothesis, we proposed a formula for calculating surprisal at the level of discourse relations. As a test case, we measured the relational surprisal according to a simple linguistic feature, i.e., presence/absence of negation in the first argument of a discourse relation.

Our analysis of the relations extracted from PDTB reveals that:

1. Sentence polarity affects the distribution of discourse relations and in particular increases the likelihood of a `chosen alternative` relation, as well as 6 other (among 44) relations. This result motivates an experimental investigation of the expectations generated by a negated sentence.

2. The sentence connective *instead* that is a strong marker of `chosen alternative` tends to be dropped in cases with lower relational surprisal, i.e., when the relation is highly predictable due to the presence of negation in Arg1. This result provides supporting evidence for UID at the level of discourse, i.e., suggesting that the writers of the WSJ text subconsciously marked instances of inter-sentential relations that are less predictable given their linguistic context.

In the future, we would like to construct a model with a high-coverage set of clausal, syntactic and semantic features predictive of relation sense classes to examine whether the relational information content of the connectives used for an instance of a relation is correlated with the predictability of the relation sense according to the features of the arguments. In a more accurate setup, it would be interesting to look into the word-by-word prediction of the model and the distance between a strong relational cue and the possibly present discourse connective.

Acknowledgments

References

Asr, F. T. and V. Demberg (2012a). Implicitness of discourse relations. In *Proceedings of the 25th International Conference on Computational Linguistics COLING*, Mumbai, India, pp. 2669 –2684.

Asr, F. T. and V. Demberg (2012b). Measuring the strength of the discourse cues. In *Proceedings of the workshop on the Advances in Discourse Analysis and its Computational Aspects*, Mumbai, India.

Asr, F. T. and V. Demberg (2013). On the information conveyed by discourse markers. In *Proceedings of the workshop on Computational Modeling and Cognitive Linguistics, ACL*, Sofia, Bulgaria.

Aylett, M. and A. Turk (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech 47*(1), 31–56.

Blakemore, D. (1992). *Understanding utterances: An introduction to pragmatics*. Blackwell Oxford.

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. In *Proceedings of the Biennial GSCL Conference*, pp. 31–40.

Buz, E., F. Jaeger, and M. K. Tanenhaus (2014). Contextual confusability leads to targeted hyperarticulation. In *Proceedings of the 36th annual conference of the cognitive science society*.

Caron, J., H. C. Micko, and M. Thuring (1988). Conjunctions and the recall of composite sentences. *Journal of Memory and Language 27*(3), 309–323.

Das, D. and M. Taboada (2013). Explicit and implicit coherence relations: A corpus study.

Demberg, V. and F. Keller (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition 109*(2), 193–210.

Demberg, V., A. B. Sayeed, P. J. Gorinski, and N. Engonopoulos (2012). Syntactic surprisal affects spoken word duration in conversational contexts. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 356–367.

Duque, E. (2013). Signaling causal coherence relations. *Discourse Studies*.

Grice, H. P. (1975). Logic and conversation. *Reprinted in Studies in the Way of Words 1985*, 2240.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research 32*(2), 101–123.

Halliday, M. and R. Hasan (1976). *Cohesion in English*. Longman (London).

Jaeger, F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology 61*(1), 23–62.

Knott, A. (1996). A data-driven methodology for motivating a set of coherence relations.

Köhne, J. and V. Demberg (2013). The time-course of processing discourse connectives. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*.

Kravtchenko, E. (2014). Predictability and syntactic production: Evidence from subject omission in russian. In *Proceedings of the 36th annual conference of the cognitive science society*.

Kurumada, C. and T. F. Jaeger (2013). Communicatively efficient language production and case-marker omission in japanese. In *The 35th Annual Meeting of the Cognitive Science Society*, pp. 858–863.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition 106*(3), 1126–1177.

Levy, R. and T. F. Jaeger (2007). Speakers optimize information density through syntactic reduction. In *Advances in Neural Information Processing Systems*.

Lin, Z., M. Kan, and H. Ng (2009). Recognizing implicit discourse relations in the penn discourse treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 343–351.

Mahowald, K., E. Fedorenko, S. T. Piantadosi, and E. Gibson (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition 126*(2), 313–318.

Millis, K. and M. Just (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*.

Murray, J. (1995). Logical connectives and local coherence. *Sources of Coherence in Reading*, 107–125.

Park, J. and C. Cardie (2012). Improving implicit discourse relation recognition through feature set optimization. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 108–112. Association for Computational Linguistics.

Patterson, G. and A. Kehler (2013). Predicting the presence of discourse connectives. In *Proceedings of the conference on Empirical Methods in Natural Language Processing EMNLP*, pp. 914–923.

Piantadosi, S. T., H. Tily, and E. Gibson (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences 108*(9), 3526–3529.

Pitler, E., A. Louis, and A. Nenkova (2009). Automatic sense prediction for implicit discourse relations in text. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pp. 683–691.

Prasad, R., N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber (2008). The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, pp. 2961–2968.

Prasad, R., A. Joshi, and B. Webber (2010). Realization of discourse relations by other means: alternative lexicalizations. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pp. 1023–1031.

Rohde, H. and W. Horton (2010). Why or what next? eye movements reveal expectations about discourse direction. In *Proceedings of 23rd Annual CUNY Conference on Human Sentence Processing*, pp. 18–20.

Rohde, H. and W. S. Horton (2014). Anticipatory looks reveal expectations about discourse relations. *Cognition 133*(3), 667–691.

Rutherford, A. T. and N. Xue (2014). Discovering implicit discourse relations through brown cluster pair representation and coreference patterns.

Sanders, T. (1997). Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes 24*(1), 119–147.

Sanders, T. J. and L. G. Noordman (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse processes 29*(1), 37–60.

Tily, H. and S. Piantadosi (2009). Refer efficiently: Use less informative expressions for more predictable meanings. In *Proceedings of the workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference*.

Trabasso, T., T. Secco, and P. Van Der Broek (1982). Causal cohesion and story coherence. *Learning and comprehension of text*.

Webber, B. (2013). What excludes an alternative in coherence relations? In *Proceedings of the International Workshop on Computational Semantics*.

Zhou, Z., Y. Xu, Z. Niu, M. Lan, J. Su, and C. Tan (2010). Predicting discourse connectives for implicit discourse relation recognition. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pp. 1507–1514.