

On the Information Conveyed by Discourse Markers

Fatemeh Torabi Asr

MMCI Cluster of Excellence
Saarland University
Germany

`fatemeh@coli.uni-saarland.de`

Vera Demberg

MMCI Cluster of Excellence
Saarland University
Germany

`vera@coli.uni-saarland.de`

Abstract

Discourse connectives play an important role in making a text coherent and helping humans to infer relations between spans of text. Using the Penn Discourse Treebank, we investigate what information relevant to inferring discourse relations is conveyed by discourse connectives, and whether the specificity of discourse relations reflects general cognitive biases for establishing coherence. We also propose an approach to measure the effect of a discourse marker on sense identification according to the different levels of a relation sense hierarchy. This will open a way to the computational modeling of discourse processing.

1 Introduction

A central question in psycholinguistic modeling is the development of models for human sentence processing difficulty. An approach that has received a lot of interest in recent years is the information-theoretic measure of *surprisal* (Hale, 2001). Recent studies have shown that surprisal can successfully account for a range of psycholinguistic effects (Levy, 2008), as well as account for effects in naturalistic broad-coverage texts (Demberg and Keller, 2008; Roark et al., 2009; Frank, 2009; Mitchell et al., 2010). : what work of Roark and Frank you mean here? Under the notion of the Uniform Information Density hypothesis (UID, Levy and Jaeger, 2007; Frank and Jaeger, 2008), surprisal has also been used to explain choices in language production: When their language gives people the option to choose between different linguistic encodings, people tend to choose the encod-

ing that distributes the information more uniformly across the sentence (where the information conveyed by a word is its surprisal).

When using surprisal as a cognitive model of processing difficulty, we hypothesize that the processing difficulty incurred by the human when processing the word is proportional to the update of the interpretation, i.e. the information conveyed by the word (Hale, 2001; Levy, 2008). We can try to estimate particular aspects of the information conveyed by a word, e.g., the information conveyed about the syntactic structure of the sentence, the semantic interpretation, or about discourse relations within the text.

This paper does not go all the way to proposing a model of discourse relation surprisal, but discusses first steps towards a model for the information conveyed by discourse connectors about discourse relations, based on available resources like the Penn Discourse Treebank (Prasad et al., 2008). First, we quantify how unambiguously specific discourse relations are marked by their typical connectors (Section 4.1) and test whether easily inferable relations are on average marked more ambiguously than relations which are less expected according to the default assumption of a reader. This idea is shaped with respect to the UID hypothesis: expected relations can afford to be signaled by weaker markers and less expected ones should be marked by strong connectors in order to keep the discourse-level information density smooth throughout the text (Section 4.2). We then investigate in more detail the types of ambiguity that a reader might face when processing discourse relations. While some ambiguities lie in discourse connectors, it also happens that more than one relation exist at the same time between two text spans. We show that

some discourse markers also signal the presence of several relations (Section 5). In computational modeling as well as laboratory setups, one should therefore have a strategy to deal with the different types of ambiguities. Finally, we ask what granularity of distinction from other discourse relations (with respect to the PDTB relation sense hierarchy) each English discourse connective conveys (Section 6).

2 Discourse Relations and their Markers

A cognitive approach to discourse processing emphasizes on the procedural role of the connectives to constrain the way readers relate the propositions in a text (Blakemore, 1992; Blass, 1993). Experimental findings suggest that these markers can facilitate the inference of specific discourse relations (Degand and Sanders, 2002), and that discourse connectors are processed incrementally Köhne and Demberg (2013). People can however infer discourse relations also in the absence of discourse connectors, relying on the propositional content of the sentences and their world-knowledge (Hobbs, 1979; Asher and Lascarides, 1998). Asr and Demberg (2012b) point out that similar inferences are also necessary for discourse relations which are only marked with a weak connector which can be used for many relations, such as *and*. Furthermore, we know that the inference of discourse relations is affected by a set of general cognitive biases. To illuminate the role of these factors let's have a look at (1). While the type of relation between the two events is clearly inferable in (1-a) and (1-b) due to the discourse connectives, in (1-c), the reader would have to access their knowledge, e.g., about Harry (from larger context) or the usual affairs between bosses and employees, in order to construct a discourse relation.

- (1) a. The boss was angry because Harry skipped the meeting (**reason**).
 b. The boss was angry, so Harry skipped the meeting (**result**).
 c. The boss was angry and Harry skipped the meeting.

Here, not only both **reason** and **result** interpretations but even an independent parallel relation (simple **Conjunction**) between the two events are possible to be inferred as a relatively

neutral connective, i.e., *and* is used. Levinson (2000) notes in his discussion on presumptive meanings that “*when events are conjoined they tend to be read as temporally successive and if at all plausible, as causally linked*”. If this is true then the **result** reading is most probable for (1-c). General preferences of this kind have been investigated via experimental approaches (Segal et al., 1991; Murray, 1997; Sanders, 2005; Kuperberg et al., 2011). Segal et al. (1991) and Murray (1997) argue that readers expect a sentence to be continuous with respect to its preceding context (the continuity hypothesis). Continuous discourse relations in terms of congruency and/or temporality are consequently easier to process than the discontinuous ones. Sanders (2005) proposes that causal relatedness entails the maximum degree of coherence in a text, therefore readers always start by attempting to find cause-consequence relations between neighboring sentences (the causality-by-default hypothesis). In a similar vein, Kuperberg et al. (2011) shows that readers face comprehension difficulty when sentences in short text spans cannot be put into causal relation and no marker of other relations (e.g., **Concession**) is available.

Taken together, these findings suggest that world knowledge, general cognitive biases, and linguistic features of the sentences such as the presence of a weak or strong marker contribute to the relational inference. With a look back to the information theoretic approach to the linguistic patterns, one could hypothesize that when one factor is strongly triggering expectation for a specific type of relation the other factors could remain silent in order to keep the information distribution uniform. With this perspective, Asr and Demberg (2012a) tested whether the predictability of discourse relations due to general cognitive biases (towards causality and continuity) can explain the presence vs. absence of the discourse connectors. They found that connectors were more likely to be dropped in the more predictable (causal or continuous) relations than in others. Our investigation of the explicit relations in this paper (the first experiment) looks into this question in a stricter manner considering **how much** information a connective delivers about discourse relations. Since this information is

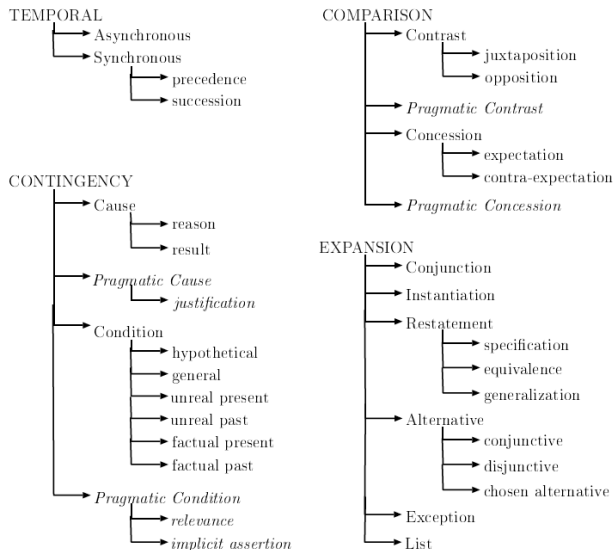


Figure 1: Hierarchy of senses in PDTB (Prasad et al., 2008)

closely related to the ambiguities a connective removes (or maybe adds to the context) in the course of reading, we dedicate a separate section in this paper to illuminate different types of ambiguities. Also, a more detail question would be **what types** of information a connective can convey about one or several discourse relations. To our best of knowledge there has been no corpus-based study so far about this last point which we will try to model in our third experiment.

3 Penn Discourse Treebank

The Penn Discourse Treebank (PDTB, Prasad et al., 2008) is a large corpus annotated with discourse relations, (covering the Wall Street Journal part of the Penn Treebank). The annotation includes sentence connectives, spans of their arguments and the sense of discourse relations implied by the connectives. The relation labels are chosen according to a hierarchy of senses (Figure 1). Annotators were asked to find the *Explicit* discourse connectives and respectively select a sense (as much specific as possible) from the hierarchy. For neighboring sentences where no explicit marker existed in the original text they were asked to first insert a suitable connective between the two arguments and then annotate a relation sense, in this case categorized as *Implicit*. If an expression — not belonging to the list of constituted connectives — in one of the involved sentences

is already indicative of a specific relation, then instead they marked that expression and put the relation into the *AltLex* category. In all of our experiments only the explicit relation are considered. Some connectives were annotated with two sense labels in the PDTB. In our analyses below, we count these text spans twice (i.e., once for each sense), resulting in a total of 19,458 relation instances.

4 Are Unexpected Relations Strongly Marked?

4.1 Markedness Measure

Point-wise mutual information (pmi) is an information-theoretic measure of association between two factors. For our purpose of measuring the markedness degree of a relation r in the corpus, we calculate the normalized pmi of it with any of the connectives, written as c that it co-occurs with:

$$\begin{aligned} npmi(r; c) &= \frac{pmi(r; c)}{-\log p(r, c)} \\ &= \frac{\log \frac{p(r, c)}{p(r)p(c)}}{-\log p(r, c)} \\ &= \frac{\log p(r)p(c)}{\log p(r, c)} - 1 \end{aligned}$$

$npmi$ is calculated in base 2 and ranges between -1 and 1 . For our markedness measure, we scale it to the interval of $[0, 1]$ and weigh it by the probability of the connector given the relation.

$$0 < \frac{npmi(r; c) + 1}{2} < 1$$

$$markedness(r) = \sum_c p(c|r) \frac{npmi(r; c) + 1}{2}$$

Intuitively, the markedness measure tells us whether a relation has very specific markers (high markedness) or whether it is usually marked by connectors that also mark many other relations (low markedness).

4.2 Discourse Expectations and Marker Strength

Given the markedness measure, we are now able to test whether those relations which are more expected given general cognitive biases

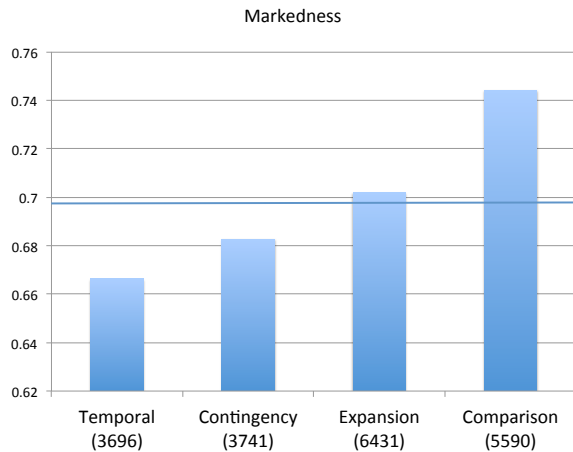


Figure 2: Markedness of level-1 explicit relations in the PDTB (frequencies of the relations given in brackets).

(expecting continuous and causal relations) are marked less strongly than e.g. discontinuous relations. Figure 2 compares the markedness associated to the explicit relations of the PDTB when the first level relation sense distinction is considered.

Figure 2 shows that **COMPARISON** relations exhibit higher markedness than other relations, meaning that discontinuity is marked with little ambiguity, i.e. markers of **COMPARISON** relations are only very rarely used in other types of discourse relations. **COMPARISON** relations are exactly those relations which were classified in Asr and Demberg (2012a) as a class of discontinuous relations. Further experimental evidence also shows that these relations are more likely to cause processing difficulty than others when no connector is present (Murray, 1997), and that their markers have a more strongly disruptive effect than other markers when used incorrectly. Under the information density view, these observations can be interpreted as markers for comparison relations causing a larger context update. The high markedness of **COMPARISON** relations is thus in line with the hypothesis that unpredictable relations are marked strongly.

CONTINGENCY relations, on the other hand, exhibit a lower score of markedness. This indeed complies with the prediction of the causality-by-default hypothesis (Sanders, 2005) in conjunction with the UID hypothesis: causal relations can still be easily inferred

even in the presence of ambiguous connectives because they are preferred by default.

As also discussed in Asr and Demberg (2012a), some types of **EXPANSION** relations are continuous while others are discontinuous; finding that the level of markedness is near the average of all relations therefore comes as no surprise.

More interesting is the case of **TEMPORAL** relations: these relations have low markedness, even though this class includes continuous (temporal succession) relations as well as discontinuous (temporal precedence) relations, and we would thus have expected a higher level of markedness than we actually find. Even when calculating markedness at the more fine-grained relation distinction level, did not find a significant difference between the markedness of the temporally forward vs. backward relations. A low level of markedness means that the connectors used to mark temporal relations are also used to mark other relations, in particular, temporal connectives are often used to mark **CONTINGENCY** relations. This observation brings us to the question of general patterns of ambiguity in discourse markers and the ambiguity of discourse relations themselves, see Section 5.

5 Ambiguous Connective vs. Ambiguous Relation

Some discourse connectives (e.g., *since*, which can be temporal or causal, or *while*, which can be temporal or contrastive) are ambiguous. In this section, we would like to distinguish between three different types of ambiguity (all with respect to the PDTB relation hierarchy):

1. A connector expressing different relations, where it is possible to say that one but not the other relation holds between the text spans, for example *since*.
2. A connector expressing a class of relations but being ambiguous with respect to the subclasses of that relation, for example *but*, which always expresses a **COMPARISON** relationship but may express any subtype of the comparison relation.
3. the ambiguity inherent in the relation between two text spans, where several rela-

Relation pair	#R1 (total)	#R2 (total)	#Pair	χ^2
T.Synchrony–CON.Cause.reason	507 (1594)	353 (1488)	187	1.08E+00
T.Asynchronous.succession–CON.Cause.reason	189 (1101)	353 (1488)	159	2.43E+02 ***
E.Conjunction–CON.Cause.result	352 (5320)	162 (752)	140	2.22E+02 ***
T.Synchrony–EXP.Conjunction	507 (1594)	352 (5320)	123	5.43E+01 ***
T.Synchrony–COM.Condition.reneral	507 (1594)	70 (362)	52	1.67E+01 ***
T.Synchrony–COM.Contrast.juxtaposition	507 (1594)	77 (1186)	45	1.97E+00
T.Asynchronous.precedence–E.Conjunction	66 (986)	352 (5320)	36	1.15E+01 ***
T.Synchrony–COM.Contrast	507 (1594)	37 (2380)	28	9.55E+00 ***
T.Synchrony–COM.Contrast.opposition	507 (1594)	28 (362)	21	6.78E+00 **

Table 1: Most frequent co-occurring relations in the PDTB, their frequency among multi-labels (and in the entire corpus)

tions can be identified to hold at the same time.

The first and second notion of ambiguity refer to what we so far have been talking about: we showed that some connectors mark can mark different types of relations, and that some connectives marking a general relation type but not marking specific subrelations.

The third type of ambiguity is also annotated in the PDTB. Relations which are ambiguous by nature are either labeled with a coarse-grained sense in the hierarchy (e.g., **COMPARISON.Contrast** the second most frequent label in the corpus chosen by the annotators when they could not agree on a more specific relation sense), or are labelled with two senses. Table 1 lists which two relation senses were most often annotated to hold at the same time in the PDTB, along with the individual frequency (also frequency in the entire corpus inside brackets). Sub-types of **Cause** and **TEMPORAL** relations appear most often together, while **TEMPORAL.Synchrony** is a label that appears significantly more than expected among the multi-label instances, even with a higher frequency than that of **EXPANSION.Conjunction**, the most frequent label in the corpus. Such observations confirm the existence of the third type of ambiguity in discourse relations.

Interestingly, these inherently ambiguous relations also have their own specific markers, such as *meanwhile* which occurs in about 70% of its instances with two relation senses¹.

¹This connective is mostly labeled with **TEMPORAL.Synchrony** and **EXPANSION.Conjunction**. Interestingly these two labels appear together significantly less frequently than expected (as marked in the table with ***) but when such a cooccurrence happened in the corpus it has been for the connective *meanwhile*.

On the other hand, other well-known ambiguous connectors like *since* rarely mark inherently ambiguous relations, and most often can be identified as one specific relation sense by looking at the content of the arguments. The importance of the possibility to annotate a second sense and hence explicitly mark the inherently ambiguous relations has also been pointed out by Versley (2011). In fact, a connective like *meanwhile* can be thought of as delivering information not only about the possible relation senses it can express, but also about the fact that two discourse relations hold simultaneously.

In conclusion, it is possible that more than one discourse relation hold between two text spans. We believe that taking into account the different types of ambiguity in discourse relations can also benefit automatic discourse relation classification methods, that so far ignore multiple relation senses. Relations with two senses mostly include one temporal sense. This also (at least partially) explains the low level of markedness of temporal relations in Figure 2. Of particular interest is also the finding that there seem to be specific connectors such as *meanwhile* which are used to mark inherently ambiguous relations.

6 Type of Information Conveyed by a Discourse Connector

In this experiment, we focus on the differences among individual connectives in reflecting information about discourse relations from coarse to fine grained granularity.

6.1 Measure of Information Gain

The mutual information between two discrete variables which is indicative of the amount of uncertainty that one removes for inference of

the other, can be decomposed in the following manner:

$$I(X; Y) = \sum_c p(c) \sum_r p(r|c) \log \frac{p(r|c)}{p(r)}$$

The inner sum is known as *Kullback-Leibler divergence* or relative entropy of the distribution of relations $p(r)$ independent of the connector c and the distribution of relations $p(r|c)$ after observing c ². The relative entropy thus quantifies in how far knowing the connector c changes the distribution of relations.

$$gain(c) = D_{KL}(p(r|c)||p(r))$$

This formulation also allows us to calculate the change in distribution for different levels of the PDTB relation sense hierarchy and thus to analyse which connectors convey information about which level of the hierarchy. We define the measure of *enhancement* to formalize this notion:

$$enhancement_{xy}(c) = gain_y(c) - gain_x(c)$$

The $enhancement_{xy}(c)$ indicates the amount of information delivered by cue c for the classification of the instances into finer-grained relation subtypes. For example, $enhancement_{01}(because)$ describes how much information gain *because* provides for distinguishing the level-1 relations it marks from other relations. Similarly, high $enhancement_{23}(because)$ indicates that this connective is important for distinguishing among level 3 relations (here, distinguishing `CONTINGENCY.Cause.reason` from `CONTINGENCY.Cause.result` relations), while low $enhancement_{23}(if)$ indicates that *if* does not contribute almost any information for distinguishing among the subtypes of the `CONTINGENCY.Condition` relation.

²Note that this formulation is closely related to surprisal: Levy (2008) shows that $surprisal(w_{k+1}) = -\log P(w_{k+1}|w_{1..w_k})$ is equivalent to the KL divergence $D(P(T|w_{1..j+1})||P(T|w_{1..j}))$ for “any stochastic generative process P , conditioned on some (possibly null) external context, that generates complete structures T , each consisting at least partly of surface strings to be identified with serial linguistic input”. Note however that in our current formulation of a discourse relation, the simplification to general structure-independent surprisal does not hold ($D_{KL}(p(r|c)||p(r)) \neq -\log p(c)$) because our relations (as they are defined here) do not satisfy the above condition for T , in particular, $P(r, c) \neq P(r)$.

6.2 Connective Help in Hierarchical Classification

Figure 3 shows the amount of enhancement for 27 frequent (> 100 occurrences) connectives in the corpus in three transitions, namely from no information to the first level classification, from first to the second level and from second to the third. Most of the connectives contribute most strongly at the coarsest level of classification, i.e., their L1-Root enhancement is the highest. In particular, we find that some of the most frequent connectives such as *but*, *and*, and *also* only help distinguishing discourse relation meaning at the coarsest level of the PDTB relation hierarchy, but contribute little to distinguish among e.g. different subtypes of `COMPARISON` or `EXPANSION`. An interesting observation is also that frequent markers of comparison relations *but*, *though*, *still* and *however* provide almost no information about the second or third level of the hierarchy.

Another group of connectors, *for example*, *instead*, *indeed* and *or* contribute significantly more information in transition from the first to the second level. These are specific markers of some level-2 relation senses. Among these, *instead* and *or* even help more for the deepest classification³.

Temporal and causal connectives such as *before*, *after*, *so*, *then*, *when* and *thus* have more contribution to the deepest classification level. This reflects the distinctions employed in the definition of the third level senses which has a direct correlation with the temporal ordering, i.e., forward vs. backward transition between the involved sentences. In other words, regardless of whatever high-level class of relation such markers fit in, the temporal information they hold make them beneficial for the 3rd level classification.

There are also a few connectives (*if*, *indeed*, *for example*) that convey a lot of information about the distinctions made at the first and second level of the hierarchy, but not about the third level. The reason for this is either that the third level distinction can only be made based on the propositional information in the

³Markers of `EXPANSION.Alternative.conjunction` and `EXPANSION.Alternative.chosen` alternative respectively.

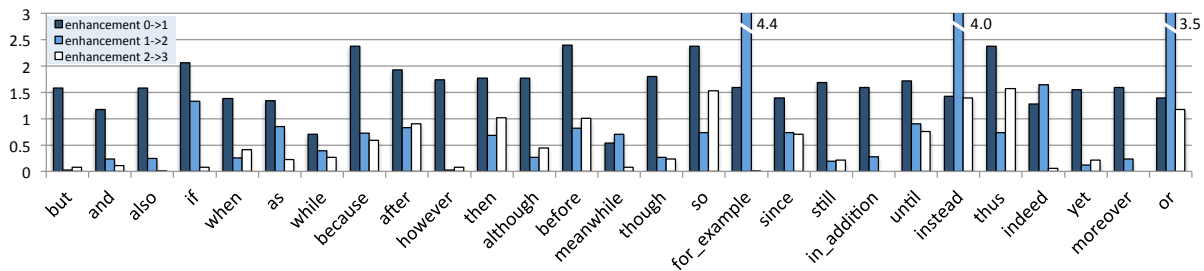


Figure 3: Enhancement through three levels of relation sense classification obtained by 27 most frequent connectives in the PDTB — ordered left to right by frequency.

arguments (this is the case for the sub-types of conditionals), or that the connector usually marks a relation which does not have a third level (e.g., *for example* is a good marker of the **EXPANSION.Instantiation** relation which does not have any subtypes).

It is worth noting that a sum over enhancements obtained in the three levels results in the total relative entropy the distribution of discourse relations before vs. after encountering the connective. As expected, ambiguous connectors of the first type of ambiguity (*while*, *since*, *when*) convey a little bit of information at each level of distinction, while overall information gain is relatively small. Ambiguous connectors of the second type of ambiguity (e.g., *but*, *and*, *if*) convey almost no information about specific sub-types of relations. Finally, markers of inherently ambiguous relations (*meanwhile*) stand out for very low information gain at all levels.

6.3 Discussion

The notion of the information conveyed by a discourse connector about a discourse relation can also help to explain two previous findings on the relative facilitative effect of causal and adversative connectors, that at first glance seem contradictory.

While Murray (1997) showed a generally more salient effect for a group of adversative cues such as *however*, *yet*, *nevertheless* and *but* compared with causal connectives *therefore*, *so*, *thus* and *consequently*, others reported different patterns when particular pairs of connectives were compared: Caron et al. (1988) found greater inference activity and recall accuracy for *because* sentences than sentences connected with *but*. Also, Millis and Just (1994) found a faster reading time and bet-

ter response to the comprehension questions in the case of *because* than that of *although* sentences. Interestingly, by looking at Figure 3, we find that *because* is a more constraining connective than *but* and even *although*, given that the information gain obtain by this connective in all levels of relation classification is greater than that of *but* and *although*. While adversative connectives are reliable signals to distinguish comparison relations in a high-level from the other three major types of relations, most causal connectives deliver specific information down to the finer grains. In particular, *because* is a distinguished marker of the **reason** relation; hence, it should be associated with a more constraining discourse effect, while a generally used connective such as *but* can serve as the marker of a variety of adversative relations, e.g., a simple **Contrast** vs. a **Concession** relation.

The information-theoretic view can also account for the larger facilitating effect of highly constraining causal and adversative connectives on discourse comprehension compared to additive connectives such as *and*, *also* and *moreover* (Murray, 1995, 1997; Ben-Anath, 2006). We also can see from the Figure 3 that the mentioned additive connectives show a relatively lower sum of enhancement.

In summary, the broad classification of a discourse connector (Murray, 1997; Halliday and Hasan, 1976) is not the only factor that determines how constraining it is, or how difficult it will be to process. Instead, one should look at its usage in different context (i.e., specificity of the connective usage in the natural text). For example, based on the measurements presented in the Figure 3 we would expect a relatively high constraining effect of the connectives such as *for example* and *instead*. Note

however that these predictions strongly depend on the discourse relation sense inventory and the discourse relation hierarchy. In particular, it is important to ask in how far computational linguistics resources, like the PDTB, reflect the inference processes in humans – in how far are the sense distinction and hierarchical classification cognitively adequate?

7 Discussion and Conclusion

Discourse Relation Hierarchy and Feature Space Dimensions Psycholinguistic models that need to be trained on annotated data from computational linguistics resources also have to be concerned about the psycholinguistic adequacy of the annotation. In particular, for a model of discourse relation surprisal, we need to ask which discourse relations are relevant to humans, and which distinctions between relations are relevant to them? For example, it may be possible that the distinction between cause and consequence (3rd level PDTB hierarchy) is more important in the inference process than the distinction between conjunction and list (2nd level PDTB hierarchy). Given the fact that more than one discourse relation (or none) can hold between two text segments, one should also ask whether a hierarchy is the right way to think about the discourse relation senses at all – it might be more adequate to think about discourse connectives conveying information about temporality, causality, contrast etc, with each connector possibly conveying information about more than one of these aspects at the same time.

These questions are also relevant for automatic discourse relation identification: many approaches to discourse relation identification have simplified the task to only distinguish between e.g. the level-1 sense distinctions, or level-2 distinctions (Versley, 2011; Lin et al., 2011; Hernault et al., 2011; Park and Cardie, 2012), but may be missing to differentiate aspects that are important also for many text interpretation tasks, such as distinguishing between causes and consequences.

Towards discourse relation surprisal A computational model of discourse relation surprisal would have to take the actual local context into account, i.e. factors other than just

the connective, and model the interplay of different factors in the arguments of the discourse relation. We would then be in a position to argue about the predictability of a specific instance of a discourse relation, as opposed to arguing based on general cognitive biases such as the causality-by-default or continuity hypotheses.

From the three studies in this paper, we note that our findings so far are compatible with a surprisal account at the discourse relation level: The first study showed that discourse relations that seem to cause a larger context update are marked by less ambiguous connectives than relations for which less information needs to be conveyed in order to be inferred. This is in line with the UID and the *continuity* and *causality-by-default* hypotheses put forth by Murray (1997) and Sanders (2005). The second study then went on to show that one can distinguish several types of ambiguity among discourse relations, in particular, more than one relation can hold between two propositions, and there are some connectives which express this inherent ambiguity. In the third study, we also showed that the effect of particular discourse markers varies with respect to their contribution in different levels of relation classification. Some connectives such as the majority of the adversative ones, simply help to distinguish contrastive relations from other classes, while those with a temporal directionality contribute most in the deeper level of the PDTB hierarchical classification. The enhancement measure introduced in this paper can be employed for measuring the effect of any discriminative feature through the hierarchical classification of the relations. This work is a first step towards the computational modeling of the discourse processing with respect to the linguistic markers of the abstract discourse relations. In future work, we would like to look at the contribution of different types of relational markers including sentence connectives, sentiment words, implicit causality verbs, negation markers, event modals etc., which in the laboratory setup have proven to affect the expectation of the readers about an upcoming discourse relation (Kehler et al., 2008; Webber, 2013).

References

- Asher, N. and Lascarides, A. (1998). Bridging. *Journal of Semantics*, 15(1):83–113.
- Asr, F. T. and Demberg, V. (2012a). Implicitness of discourse relations. In *Proceedings of COLING*, Mumbai, India.
- Asr, F. T. and Demberg, V. (2012b). Measuring the strength of the discourse cues. In *workshop on the Advances in Discourse Analysis and its Computational Aspects*, Mumbai, India.
- Ben-Anath, D. (2006). The role of connectives in text comprehension. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 5(2).
- Blakemore, D. (1992). *Understanding utterances: An introduction to pragmatics*. Blackwell Oxford.
- Blass, R. (1993). Are there logical relations in a text? *Lingua*, 90(1-2):91–110.
- Caron, J., Micko, H. C., and Thuring, M. (1988). Conjunctions and the recall of composite sentences. *Journal of Memory and Language*, 27(3):309–323.
- Degand, L. and Sanders, T. (2002). The impact of relational markers on expository text comprehension in l1 and l2. *Reading and Writing*, 15(7):739–757.
- Demberg, V. and Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210.
- Frank, A. and Jaeger, T. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. *Proceedings of the 28th meeting of the Cognitive Science Society*.
- Frank, S. (2009). Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In *Proceedings of the 31st annual conference of the cognitive science society*, pages 1139–1144.
- Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001*, pages 1–8.
- Halliday, M. and Hasan, R. (1976). *Cohesion in English*. Longman (London).
- Hernault, H., Bollegala, D., and Ishizuka, M. (2011). Semi-supervised discourse relation classification with structural learning. *Computational Linguistics and Intelligent Text Processing*, pages 340–352.
- Hobbs, J. R. (1979). Coherence and coreference. *Cognitive science*, 3(1):67–90.
- Kehler, A., Kertz, L., Rohde, H., and Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25(1):1–44.
- Köhne, J. and Demberg, V. (2013). The time-course of processing discourse connectives. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*.
- Kuperberg, G., Paczynski, M., and Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, 23(5):1230–1246.
- Levinson, S. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. The MIT Press.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.
- Levy, R. and Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In *Advances in Neural Information Processing Systems*.
- Lin, Z., Ng, H., and Kan, M. (2011). Automatically evaluating text coherence using discourse relations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 997–1006.
- Millis, K. and Just, M. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*.
- Mitchell, J., Lapata, M., Demberg, V., and Keller, F. (2010). Syntactic and semantic factors in processing difficulty: An integrated measure. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 196–206.
- Murray, J. (1995). Logical connectives and local coherence. *Sources of Coherence in Reading*, pages 107–125.

- Murray, J. (1997). Connectives and narrative text: The role of continuity. *Memory and Cognition*, 25(2):227–236.
- Park, J. and Cardie, C. (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*, pages 108–112. Association for Computational Linguistics.
- 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.
- Roark, B., Bachrach, A., Cardenas, C., and Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 324–333, Singapore. Association for Computational Linguistics.
- Sanders, T. (2005). Coherence, causality and cognitive complexity in discourse. In *Proceedings/Actes SEM-05, First International Symposium on the Exploration and Modelling of Meaning*, pages 105–114.
- Segal, E., Duchan, J., and Scott, P. (1991). The role of interclausal connectives in narrative structuring: Evidence from adults' interpretations of simple stories. *Discourse Processes*, 14(1):27–54.
- Versley, Y. (2011). Towards finer-grained tagging of discourse connectives. In *Proceedings of the Workshop Beyond Semantics: Corpus-based Investigations of Pragmatic and Discourse Phenomena*.
- Webber, B. (2013). What excludes an alternative in coherence relations? In *Proceedings of the IWCS*.