The Overall Markedness of Discourse Relations

Lifeng Jin and Marie-Catherine de Marneffe
Department of Linguistics
The Ohio State University
{jin, mcdm}@ling.osu.edu

Abstract

Discourse relations can be categorized as continuous or discontinuous in the hypothesis of continuity (Murray, 1997), with continuous relations expressing normal succession of events in discourse such as temporal, spatial or causal. Asr and Demberg (2013) propose a markedness measure to test the prediction that discontinuous relations may have more unambiguous connectives, but restrict the markedness calculation to relations with explicit connectives only. This paper extends their measure to explicit and implicit relations and shows that results from this extension better fit the continuity hypothesis predictions both for the English Penn Discourse (Prasad et al., 2008) and the Chinese Discourse (Zhou and Xue, 2015) Treebanks.

1 Introduction

Discourse relations between units of text are crucial for the production and understanding of discourse. Different taxonomies of discourse relations have been proposed (i.a. Hobbs (1985), Lascarides and Asher (1993) and Knott and Sanders (1998)). One taxonomy is based on deictic continuity (Segal et al., 1991; Murray, 1997): continuity in the sense of Segal et al. (1991) means that the same frame of reference is maintained, for example by subsequent sentences talking about the same event, without a shift in perspective (Asr and Demberg, 2012). For instance, a causal relation such as *I was tired, so I drank a cup of coffee.* is continuous, and adversatives show discontinuous relations: *I drank a cup of coffee but I was still tired.* Other continuous relations include temporal succession, topic succession and so on. The continuity hypothesis predicts that sentences connected by continuous relations are easier to understand than ones connected by discontinuous relations.

Previous work on continuity hypothesis (Maury and Teisserenc, 2005; Cain and Nash, 2011; Hoek and Zufferey, 2015) suggests that discourse connectives are indicators of the continuity of discourse and help the interlocutors predict the level of continuity of upcoming sentences. Segal et al. (1991) propose that connectives which signal discontinuous discourse relations, such as *but*, are the marked ones because they indicate harder-to-comprehend content. Asr and Demberg (2012, 2013) extend this idea to discourse relations, proposing that discourse relations which are discontinuous, or posing a conceptual difficulty (Haspelmath, 2006), may be less explicitly conveyed in text, or more explicitly marked by a connective which unambiguously conveys that specific relation than continuous ones. They propose a new measure called markedness to capture this, but when computed on the Penn Discourse Treebank, results do not fit the continuity theory well. This paper improves on Asr and Demberg (2013)'s measure and shows that the results on the Penn Discourse and the Chinese Discourse Treebanks fit the continuity hypothesis very well.

2 Discourse Treebanks

Penn Discourse Treebank The Penn Discourse Treebank (PDTB) is a corpus of Wall Street Journal articles annotated with discourse relations (Prasad et al., 2008). The discourse relations are organized in a hierarchical structure with three levels: a level 1 (e.g. TEMPORAL), level 1/level 2 (e.g. TEMPORAL.Asynchronous) or level 1/level 2/level 3 (e.g. TEMPORAL.Asynchronous.succession) relation can appear between two clauses within a sentence. Discourse relations with overt discourse connectives are annotated as "Explicit", whereas the relations with no discourse connective are annotated as "Implicit". The "AltLex" category is used when a non-connective expression conveys the relation. Table 1 gives the distribution of the

relation categories in the corpus.

Some connectives are labeled with multiple relations when it was difficult to pinpoint exactly one exact discourse relation for it. We follow Asr and Demberg (2013) and treat these cases as if there are multiple instances of the connective, each with one of the labels it received. This gives us a total of 35,870 relation instances.

	PDTB		CDTB	
Relation Category	**Count**	**%**	**Count**	**%**
Explicit	18459	45.5	1219	22.0
Implicit	16053	39.5	3935	71.1
AltLex	633	1.6	116	2.1
EntRel and NoRel	5464	13.4	264	4.8

Table 1: Relation category counts for the Penn Discourse and Chinese Discourse Treebanks.

Chinese Discourse Treebank The Chinese Discourse Treebank (CDTB, Zhou and Xue 2014) follows the PDTB annotation style and has annotations for 164 documents from Xinhua News. The main difference between CDTB and PDTB is that CDTB has a flat structure of only ten relations compared to the hierarchical relation structure in PDTB. Table 1 gives the distribution of the relation categories.

3 Rethinking the Markedness Measure

To quantify the conceptual difficulty of discourse relations, Asr and Demberg (2013) propose an information-theoretic measure "markedness", which tells us how tightly and uniquely a relation is associated with a connective. The measure uses normalized point-wise mutual information:

$$npmi(r; c) = \frac{\log p(r)p(c)}{\log p(r, c)} - 1 \qquad (1)$$

to get the markedness of a discourse relation:

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

where r is a relation and c is a discourse connective. Asr and Demberg (2013) propose this measure in the surprisal framework of Levy (2008), and restrict the scope of the data to only Explicit relations in PDTB. We will call this measure with only explicit relations "M-exp". Since surprisal is defined as the probability of a word given previous

words and context (3), this restriction on the scope of relations does not seem reasonable. Surprisal is defined as

$$surprisal \propto - \log p(w_i | w_{1...i-1}, CONTEXT) \qquad (3)$$

It can be argued that at the current word w_{i-1}, the distribution of upcoming discourse relations, available in CONTEXT, should play a role in determining the probability of the upcoming word w_i. In the surprisal model, the domain for w_i should be the same as w_{i-1}, which is all possible words. However, if we only calculate the distribution of explicit connectives as proposed by Asr and Demberg (2013), the candidates for w_i will change according to the prediction of whether an explicit relation is coming up or not. If the upcoming relation is an implicit relation, one then has access to a distribution of words without the connectives, whereas if one predicts an explicit relation, then one predicts the next word using a distribution of all the connectives as in M-exp. However surprisal should not be a model of deterministic decision making. It is more likely the case that given CONTEXT, one assigns probabilities to all words given the preceding context, which includes the case where no connective, in other words a zero or null connective, is predicted. The null connective may also be viewed as the probability mass for all the non-connective words predicted by CONTEXT where the connective is predicted to be omitted (Asr and Demberg, 2015).

The markedness measure can be analyzed in terms of point-wise mutual information (*pmi*), indicating the amount of information one relation has for the distribution of the words that follow it (Hume, 2011). Because *pmi* is proportionate to *npmi*, we can rewrite (2) as

$$markedness(r) \propto \sum_c p(c|r) pmi(r; c) \qquad (4)$$

We also have the mutual information measure:

$$I(X; Y) = \sum_y \sum_x p(x, y) pmi(x; y) \qquad (5)$$

For the mutual information of y_i in Y :

$$\begin{aligned} I(X; Y = y_i) &= \frac{\sum_x p(x, y_i) pmi(x; y_i)}{p(y_i)} \\ &= D_{KL}(p(x|y_i) \parallel p(x)) \\ &= \sum_x p(x|y_i) pmi(x; y_i) \\ &\propto markedness(y_i) \end{aligned} \qquad (6)$$

Level 1 Relation	Level 2 Relation	Level 3 Relation	Continuity
TEMPORAL	Asynchronous	precedence	Discontinuous
		succession	Discontinuous
	Synchronous		Ambiguous
CONTIGENCY	Cause		Continuous
	Condition		Unidentified
COMPARISON			Discontinuous
EXPANSION	Instantiation		Continuous
	Restatement		Continuous
	List		Continuous
	Alternative		Discontinuous
	Exception		Discontinuous
	Conjunction		Ambiguous

Table 2: Continuity of relations according to the continuity hypothesis (Asr and Demberg, 2012).

Therefore, the markedness measure can be understood as the Kullback-Leibler divergence of the univariate distribution of X from the conditional distribution of X given y_i (a discourse relation in our case). This shows the influence a relation has on the unigram distribution of words. Discourse relations strongly associated with certain connectives will have larger values of this measure than the ones with a weak association. In the previous discussion of surprisal, we have seen that one may treat the implicit cases as predicting a null connective, which will then expand the domain of X from explicit cases to all implicit and explicit cases. With this setup, we calculate "M-all" using all explicit and implicit relations, with a null connective accompanying all the implicit relations.

Continuity hypothesis has been linked with cognitive difficulties in discourse processing in previous studies (Segal et al., 1991; Murray, 1997). The markedness measure can also be linked to processing difficulties through surprisal theory. Surprisal theory proposes that processing difficulty during sentence processing can be seen as the work incurred by resource allocation during parallel disambiguation (Levy, 2008). If a relation has a high markedness value, it indicates that this relation has a strong influence on the distribution of upcoming candidate words. The stronger the influence is, the higher the resource allocation cost will be for the relation, thus more difficult to process.

4 Results on PDTB

Figure 1 compares the markedness of the PDTB level 1 relations as computed by the M-exp and M-all measures. According to the continuity hypothesis, the TEMPORAL relation is discontinu-
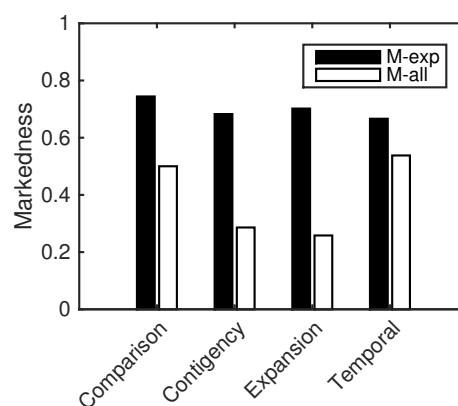

Figure 1: Comparison of markedness measures for PDTB level 1 relations.

ous (see Table 2 which gives the classification of discourse relations according to the continuity hypothesis). It should therefore have a high markedness value. However, M-exp assigns a low markedness to TEMPORAL, which Asr and Demberg (2013) note is unexpected. They ascribe this to the fact that temporal discourse connectives are often used to mark CONTINGENCY relations. However because of the high counts of Explicit connectives in TEMPORAL, whenever there is a connective that can indicate CONTINGENCY or TEMPORAL, one is more likely to predict TEMPORAL because of fewer null connective cases for TEMPORAL. In surprisal terms, whenever one predicts that there is a TEMPORAL relation next, one will more likely predict that there is an explicit discourse connective signaling the relation.

EXPANSION is the least marked of all the relations in Figure 1 with M-all. An analysis of the level 2 relations explains this fact. Figure 2 com-
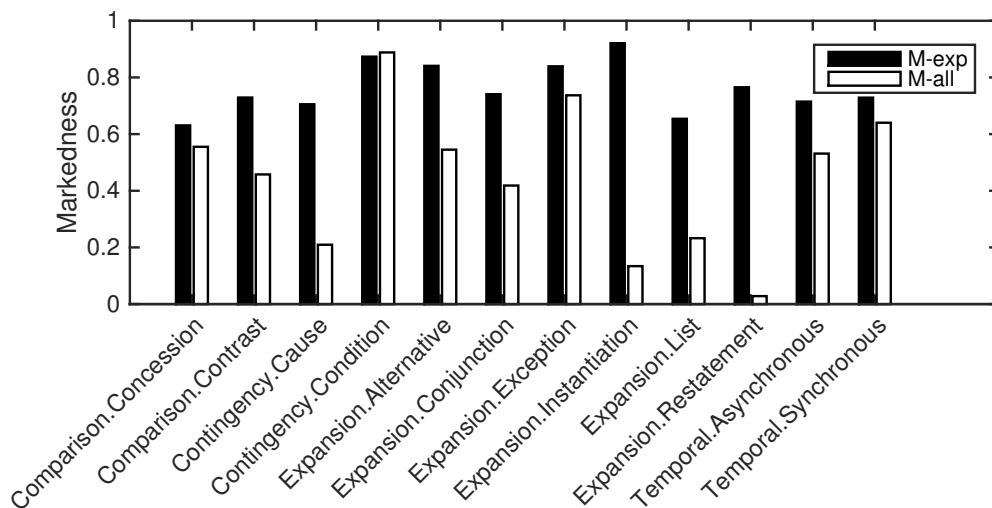
1116

Figure 2: Comparison of markedness measures for PDTB level 2 relations.

pares both measures for the level 2 relations.[1] Using M-all, it is easy to see that discontinuous relations and ambiguous relations are more marked than the continuous ones. In the case of EXPANSION, the level 2 continuous relations are among the least marked ones, which are keeping the overall markedness low. Also, the ones which are discontinuous, especially *Exception*, are rare, so their influence to the overall score for EXPANSION is small. The most frequent relation, *Conjunction*, can be viewed as sometimes continuous and sometimes discontinuous, therefore the overall markedness rating for it is in the middle. All these factors contribute to the lowest markedness for EXPANSION. For CONTINGENCY, *Condition* is not classified as continuous or discontinuous, and it is highly marked, thus driving the overall score high. However if *Condition* is removed, then CONTINGENCY will be the least marked relation at the level 1.

At the level 3, there are two discontinuous relations of interest: *precedence* (e.g. *I had a cup of coffee before I took a bath*) and *succession* (e.g. *I took a bath after I had a cup of coffee*) under the TEMPORAL.Asynchronous relation. Table 3 compares the markedness measures for both relations. Asr and Demberg (2013) mention that there is no significant markedness distinction for them and we can see that *precedence* is slightly more marked than *succession* in M-exp, but the differ-

Metric	Precedence	Succession
M-exp	0.799	0.783
M-all	0.494	0.687

Table 3: Precedence and Succession Markedness.

	Arg1-Conn-Arg2		Conn-Arg2-Arg1	
	Implicit	Explicit	Implicit	Explicit
Precedence	567	931	0	55
Succession	171	867	0	234

Table 4: Counts for different argument orders of Precedence and Succession.

ence is small. M-all however shows that *succession* is more marked than *precedence*, reflecting the fact that *precedence* is easier to understand.

For *precedence*, the arguments can be in a normal temporal order, i.e. a forward temporal order (Arg1-Conn-Arg2, e.g. *I had some coffee before I went out.*) or in a backward temporal order (Conn-Arg2-Arg1, e.g. *Before I went out, I had some coffee.*). For *succession*, the temporal order and the argument order are reversed. Table 4 gives the counts in PTDB for both *precedence* and *succession* with different argument orders, showing that Arg1-Conn-Arg2 is the most frequent construction for both relations, which is forward temporal order for *precedence* and backward temporal order for *succession*, despite the fact that both of the relations can follow a forward temporal order. Therefore the results from M-all match the continuity hypothesis prediction that events in for-

---

[1]Pragmatic relations are not shown due to their small number of occurrences.

ward temporal order are easier to understand, thus less marked, than events in backward temporal order. Asr and Demberg (2012) explain that the relatively high count of Conn-Arg2-Arg1 constructions in *succession* is due to the fact that this construction actually places the events in the forward temporal order. We also notice that *precedence* has a lot more implicit occurrences than *succession*, meaning that inferring a normal temporal relation is much easier than inferring a reversed temporal relation.
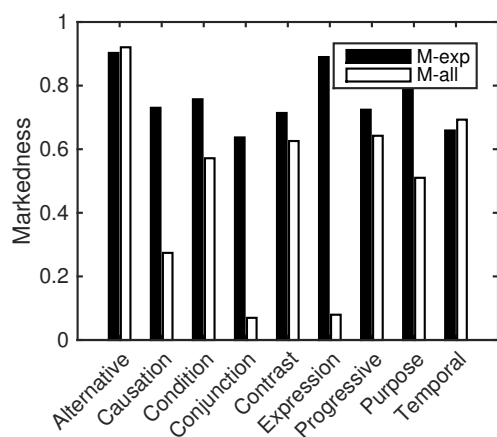
5 Results on CDTB and Comparison

Figure 3: Markedness comparison for CDTB.

The results for CDTB, in Figure 3, shows the same trends as for English: overall, the markedness computed by M-all better fits the continuity hypothesis. EXPANSION is considered by M-exp as the second highest marked relation, whereas the continuity hypothesis predicts it to be one of the lowest marked relation, which is correctly captured by M-all. The reason for it to be low may be that discontinuous relations included by EXPANSION are rare so the frequent continuous relations dominate, just as for English. Using M-exp, CAUSATION is around the middle, but M-all correctly lowers it to the third least marked relation. TEMPORAL is now the second highest marked relation among all relations, as opposed to the second least marked one. Importantly, M-all correctly shows that discourses relations in English and Chinese behave similarly in terms of markedness, which indicates that the continuity hypothesis is valid across languages.

Sanders (2005) proposes the causality-by-default hypothesis, claiming that CAUSATION is

the default discourse relation when processing discourse. However, looking at M-all scores for both languages, it is clear that CAUSATION is not the least marked relation in either language. In fact, EXPANSION can actually be seen as the least marked common relation in both languages, which may indicate that EXPANSION is the default discourse relation cross-linguistically, yet more investigations are needed to decide which one in EXPANSION is the default. CONJUNCTION is also among the least marked relations in Chinese, with 89% of its instances being implicit, but in English, CONJUNCTION has an average markedness score. This shows that there are also differences among languages on judgments of continuity of specific relations.

6 Conclusion

The continuity hypothesis predicts that discontinuous discourse relations are less expected than continuous relations and should be more marked. We expand Asr and Demberg (2013)'s measure from explicit relations only to explicit and implicit relations. We show that the results from that expansion fit the predictions of the theory very well, and such evidence demonstrates that discontinuous relations are indeed cognitively more difficult to process. Further we show that such difficulty is consistent across languages, indicating that discourse relations may not be influenced by idiosyncrasies of specific languages. Apart from the markedness measure, Asr and Demberg (2012) proposed an "implicitness" measure, modeling the continuity of a relation using the ratio of implicit cases to all cases. Incorporating explicit and implicit relations into the markedness measure has the advantage of not only providing a single measure but also one which better fits the continuity hypothesis and surprisal theory.

References

Fatemeh Torabi Asr and Vera Demberg. 2012. Implicitness of discourse relations. In *Proceedings of COLING 2012*, pages 2669–2684.

Fatemeh Torabi Asr and Vera Demberg. 2013. On

the information conveyed by discourse markers. In *Proceedings of the Fourth Annual Workshop on Cognitive Modeling and Computational Linguistics (CMCL)*, pages 84–93. Association for Computational Linguistics.

Fatemeh Torabi Asr and Vera Demberg. 2015. Uniform surprisal at the level of discourse relations: negation markers and discourse connective omission. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 118–128. Association for Computational Linguistics.

Kate Cain and Hannah M. Nash. 2011. The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology*, 103(2):429–441.

Martin Haspelmath. 2006. Against markedness (and what to replace it with). *Journal of Linguistics*, 42:25–70, 3.

Jerry R Hobbs. 1985. *On the coherence and structure of discourse*. CSLI.

Jet Hoek and Sandrine Zufferey. 2015. Factors influencing the implicitation of discourse relations across languages. In *Proceedings 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (isa-11)*, page 39.

Elizabeth Hume. 2011. Markedness. In Marc van Oostendrop, Colin J. Ewen, Elizabeth Hume, and Keren Rice, editors, *The Blackwell Companion to Phonology: Suprasegmental and prosodic phonology*, volume 2, pages 79–106. John Wiley & Sons.

Alistair Knott and Ted J.M. Sanders. 1998. The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics*, 30:135–175.

Alex Lascarides and Nicholas Asher. 1993. Temporal interpretation, discourse relations, and common sense entailment. *Linguistics and Philosophy*, 16:437–493.

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

Pascale Maury and Amélie Teisserenc. 2005. The role of connectives in science text comprehension and memory. *Language and Cognitive Processes*, 20(3):489–512.

John D. Murray. 1997. Connectives and narrative text: The role of continuity. *Memory & Cognition*, 25(2):227–236.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of LREC 2008*.

Ted Sanders. 2005. Coherence, causality and cognitive complexity in discourse. In *Proceedings of the Symposium on the Exploration and Modelling of Meaning*.

Erwin M. Segal, Judith F. Duchan, and Paula J. Scott. 1991. The role of interclausal connectives in narrative structuring: Evidence from adults' interpretations of simple stories. *Discourse Processes*, 14(1):27–54.

Yuping Zhou and Nianwen Xue. 2015. The Chinese Discourse TreeBank: a Chinese corpus annotated with discourse relations. *Language Resources and Evaluation*, 49(2):397–431.