

Can Discourse Relations be Identified Incrementally?

Frances Yung

Language Science and Technology,
Saarland University,
Saarland Informatic Campus,
66123 Saarbrücken, Germany
frances@coli.uni-saarland.de

Hiroshi Noji Yuji Matsumoto

Information Science,
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma
Nara, 630-0101, Japan
{noji,matsu}@is.naist.jp

Abstract

Humans process language word by word and construct partial linguistic structures on the fly before the end of the sentence is perceived. Inspired by this cognitive ability, incremental algorithms for natural language processing tasks have been proposed and demonstrated promising performance. For discourse relation (DR) parsing, however, it is not yet clear to what extent humans can recognize DRs incrementally, because the latent ‘nodes’ of discourse structure can span clauses and sentences. To answer this question, this work investigates incrementality in discourse processing based on a corpus annotated with DR signals. We find that DRs are dominantly signaled at the boundary between the two constituent discourse units. The findings complement existing psycholinguistic theories on expectation in discourse processing and provide direction for incremental discourse parsing.

1 Introduction

Incremental processing is an essential characteristic of human language comprehension, because linguistic data naturally occurs in streams. For example, during sentence comprehension, humans do not start parsing only after the whole sentence is perceived. Instead the human processor incrementally constructs a partial syntactic tree that matches the sentence prefix read so far (Tanenhaus et al., 1995). Though intuitively for parsing every word is relevant to the syntactical structure, it may not be the case for more global linguistic structures such as DRs, which may only be triggered by some informative cue words, and it is yet unclear at which point the human processor recog-

nizes a DR as the sentence is read or listened word by word.

DRs are relations between units of texts, such as clauses or sentences. For example,

1. In the first year, the bank eliminated 800 jobs. Now it says it will trim more in the next year.

the first and second sentences are connected by a *temporal* relation, as the events occur in a temporal sequence. An incremental discourse processor should predict and recognize the relation at some point before the end of the second sentence. It is useful for speech recognition and dialogue systems where real-time analysis is desirable.

Towards an incremental approach to automatic discourse parsing, this work investigates word-level incrementality in human discourse processing based on manual identification of DR cues. The point at which humans recognize a DR can provide a reference on how long an incremental discourse parser should ‘wait’, before there is enough input for timely yet accurate prediction of the discourse structure.

2 Related Work

This work is related to incremental approaches of natural language processing (NLP) and psycholinguistic studies on human discourse processing.

In NLP, incremental approaches are used in tasks such as syntactic parsing (Stolcke, 1995; Collins and Roark, 2004; Köhn and Menzel, 2014), semantic role labeling (Konstas et al., 2014) and other joint tasks (Hatori et al., 2012; Li and Ji, 2014; Zhou et al., 2016). These incremental systems are advantageous since they are capable of synchronous analysis by accepting sentence prefixes as inputs. On top of generating more natural and timely response in dialogue systems and improving language modeling in speech recognition,

these models can also be used to reflect difficulties in human language processing (Keller, 2010; Demberg et al., 2013).

However, we are not aware of any prior work that implements a discourse processor with such a strong assumption to incrementality. Although expectation for upcoming DRs is demonstrated in various lexico-syntactic constructions in the first clause/sentence (Cristea and Webber, 1997), existing methods of discourse parsing rely on a pipeline, in which the raw text is first segmented into discourse units, mostly clauses or sentences, and the relation is predicted based on two complete discourse units. In this respect, even shift-reduce discourse parsers (Marcu, 1999; Reitter, 2003; Sagae, 2009; Ji and Eisenstein, 2014) are incremental only at discourse unit level.

In psycholinguistics, expectation in language processing is a well studied topic (e.g. Altmann (1998)). Experimental studies suggest that humans use available pragmatic cues to generate expectations and anticipate the upcoming discourse structure (Rohde, 2008), but there are diverging findings about the time-course for humans to recognize and integrate DRs. For example, Millis and Just (1994) state that integration of a *causal* relation takes place at the end of the second clause. In contrast, other experiments report that the integration already occurs in the beginning of the second clause, at least for some relation types (Traxler et al., 1997; Cozijn, 2000; Mak and Sanders, 2010, 2013; Köhne and Demberg, 2013). These experiments are, however, limited to comparison of a few relation types and mostly depend on discourse markers (e.g. *however*, *because*). We still lack an integrated picture on where humans generally recognize a DR.

3 Methodology

This study presents an off-line corpus analysis to determine when or where humans recognize a DR as they process words incrementally. To this end, we want a human subject to identify the cues within the component clauses/sentences that trigger the recognition of a given DR, such as the underlined tokens in Example (1).

Although the exact annotated resource is not yet available, we obtained such annotation by converting the annotation in the RST Signaling Corpus (Das et al., 2015).

Data The RST Signaling Corpus consists of annotation of discourse signals over the RST Discourse Treebank (Carlson et al., 2002), which is a discourse annotated resource following the Rhetorical Structure Theory (RTS) (Mann and Thompson, 1988). In the RST Discourse Treebank, a DR is annotated between two consecutive discourse units. In turn, in the RST Signaling Corpus, each DR is further labeled with one or more types of signaling strategy. These signals not only include explicit discourse markers but also other features typically used in automatic implicit relation identification and psycholinguistic research, such as reference, lexical, semantic, syntactic, graphical and genre features (Das and Taboada, 2017). For example, the *temporal* relation in Example (A) is annotated with three signal labels in the RST Signaling Corpus:¹

- (1) discourse marker (*now*)
- (2) tense (*past — present, future*)
- (3) lexical chain (*first year — next year*)

Only 7% of the relations are annotated as ‘implicit’. Therefore, most conventionally ‘implicit’ relations are also annotated with explicit signals and included in the present analysis.

Locating signal positions Based on these labels, we use heuristic rules (see *appendix*) and gold syntactic annotation² to identify the actual cue words in the text. For example, based on the above 3 signal labels, we identify the underlined tokens in Example (1). Manual check on 200 random samples shows that all signal tokens are perfectly tagged in 95% of the samples, and the remaining 5% samples are partially correct.

We focus on relations that are signaled by surface tokens in order to examine word-level incrementality in discourse processing. Thus, we do not consider signals that are not associated with particular words, e.g. *genre*, and relations with annotations that are not specific enough. 4,146 relations are screened³ and 15,977 relations are included in the analysis. The distribution of the DRs under analysis is shown in Table 1.

¹The list of DR signals and the relation between the RST Treebank and the RST Signaling Corpus can be found in the *appendix*. Details can be found in the related literature.

²provided by the Penn Treebank, which annotates on the same text as the RST Treebank (Marcus et al., 1993)

³List of excluded signals are shown in the appendix.

category	relation sense	count
expansion	elaboration	7,070
	joint	1,031
	background	787
	evaluation	505
	manner-mean	197
	summary	170
	topic-comment	44
	topic-change	21
comparison	contrast	934
	comparison	243
contingency	enablement	512
	cause	499
	explanation	325
	condition	263
temporal	temporal	429
attribution	attribution	2,947
Total		15,997

Table 1: Sense distribution of discourse samples used in the analysis. The original RST senses are mapped to 18 conventional senses (2 screened)

Relating signal positions to incremental processing We analyze the positions of the cue tokens in relation to the DRs they signal. Each cue position is represented by its *distance* from the *boundary* of the relation’s discourse units. The *boundary* is defined as the first word of the second clause/sentence in the relation, as each relation is annotated between two consecutive clauses/sentences in the RST formalism.⁴ For example, the cue words *eliminated* and *now* in Example (1) have distances of -4 and 0 , respectively.

Although positions of the discourse cues can be identified from the recovered annotation, it is still unclear how informative the discourse cues are. It is possible that unambiguous cues only occur at the end even though numerous cues occur in the beginning. For example, in Example 1, can people correctly anticipate the temporal relation after reading the word *now*? Or is *now* too ambiguous that it is necessary to consider all signals after reading the last word? To answer these questions, we quantify and compare the

⁴ Some relations, e.g. *list*, have more than two consecutive units. In this case, the *distance* of the cue is the *distance* compared with the closest *boundary*.

discourse informativeness of prefixes in different sizes.

The informativeness of each prefix is calculated from the cues covered by the prefix. For each DR spanning two consecutive clauses/sentences, the prefix size ranges from the first word of the first clause/sentence to the complete first and second clauses/sentences. Consecutive cue tokens are merged as one signal and a signal is counted as being covered by a prefix only if the last token of the signal string⁵ is covered by the prefix. We use majority as a baseline approach to associate the discourse signals with the relation sense. The inferred relation sense r_{p_n} based on the majority cues in discourse prefix p_n is defined as:

$$r_{p_n} = \arg \max_{r \in R} \sum_{s \in S_{p_n}} \text{count}(s, r) \quad (1)$$

where R is the set of all relation senses; S_{p_n} is the set of signal strings covered in discourse prefix p_n ; n is the *distance* of the last word of p_n ; and $\text{count}(s, r)$ is the count of string s being identified as a signal for a DR of sense r in the corpus. The most frequent relation, *elaboration*, is assigned if no signals are found in the prefix.

The relation senses inferred from prefixes of various sizes are compared with the actual relation sense. Although the majority approach does not model inter-relation and ambiguity of the signals, we assume that more signals, and thus longer prefixes, give better or the same prediction⁶. Therefore, we can compare the informativeness of the prefixes with that of the whole discourse span as upper bound.

4 Results

Distribution of signal locations This analysis seeks to find out how far humans read before they recognize a DR. If DR cues are evenly distributed throughout the discourse components, partial discourse structures can plausibly be constructed on the fly. On the other hand, if the relation cues generally occur towards the end of the last clause, integration of the DR is better to be restrained until

⁵ or the last token of the last span if the signal has multiple spans of strings, such as *first year – next year*

⁶ Empirically, this assumption was true: in over 99% of the relation samples, majority prediction based on signals in both clauses is better or the same, as that based on the first clause alone.

all clauses are perceived, implying limited incrementality in discourse processing.

Result of the analysis reveals that it is neither of the cases. Figure 1 shows the relative *distance* of the signals with respect to the length of the discourse units. It can be observed that most signals occur at the boundary, and the further away from the boundary, the less signals are found. In fact 24% of the tagged tokens belong to the first 2 words of the second discourse unit. Note that these do not limit to explicit discourse connectives but also other lexical and semantic features.

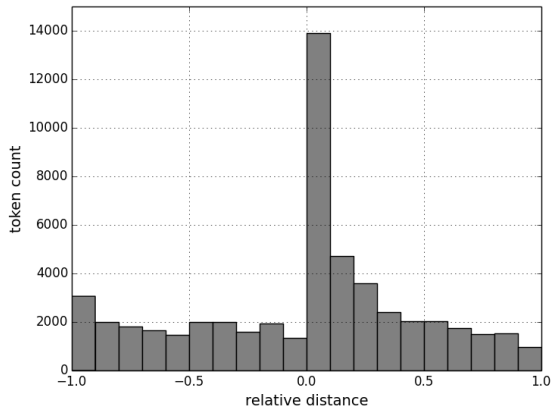


Figure 1: Distribution of the relative distances of the signal tokens.

Overall, more signal tokens locate after the boundary. Counting by relation, 52% of the relations have signals only in the second discourse unit (49% of which at the boundary), 20% have signals only in the first discourse unit, and 28% have signals in both. In other words, in 69% of the cases, all signals for the DR are covered after reading the relation boundary.

Informativeness of discourse prefixes Similarly, the informativeness of the discourse prefixes shows a leap across the *boundary*. Figure 2 illustrates the accuracy of the DR predicted by prefixes of all the relation samples collectively. *Accuracy* refers to the proportion that r_{p_n} equals the actual relation sense of the discourse sample. The upper half of Figure 2 shows that the prediction *accuracy* rises sharply after the boundary is read. According to Figure 1, more signals are detected in the first clause near the boundary, but the informativeness of the prefixes actually drops slightly, possibly due to the ambiguity of the signals. Yet the drop is reverted at the boundary and the accuracy

remains stable. This implies that the signals occurring later in the second clause do not contradict to those found at the boundary.

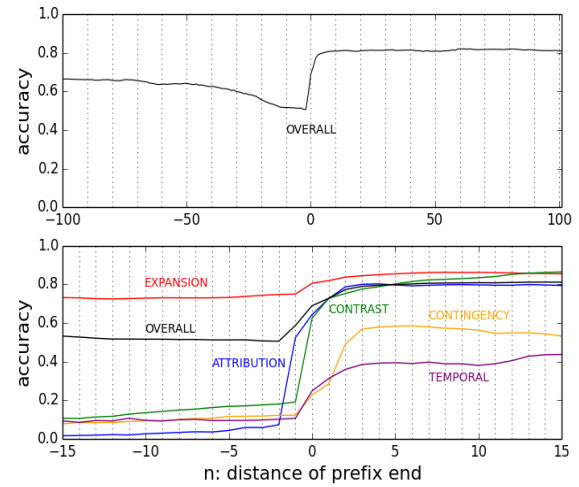


Figure 2: ‘Accuracy’ of sense prediction based on oracle signals covered by discourse prefixes⁸.

The lower half of Figure 2 compares the five sense categories defined in Table 1, zooming at prefixes ending near the *boundary*. It is observed that, in general, signals for *contingency* and *temporal* relations are mostly identified just after the boundary, while *expansion*, *attribution* and *contrast* relations are identified just before the boundary. The informativeness of the discourse prefixes of *expansion* relation does not rise sharply like other relations because it is the default relation when no signals are identified. Nonetheless, it still hold for all relations that predictions just after the boundary is similar to predictions at the end of the second discourse unit.

5 Conclusion

This work investigates whether DRs can be identified incrementally based on human performance. Our analysis concludes that it is possible because DR signals occur throughout the discourse. Nonetheless, the signals are not evenly distributed but concentrated on the boundary of the two discourse units. An incremental discourse parser that jointly segments discourse units and predicts DR senses can potentially output the predicted DR immediately after a *boundary* is detected, and then

⁸This ‘accuracy’ is not comparable to the performance of automatic parsers because the signals are identified manually and the prediction is not made on a held out test set. Our focus is the comparison between the discourse prefixes.

focus on detecting *expectative* signals in the second clause/sentence for the next relation.

Results of the analysis agree with the psycholinguistics literature that DRs are integrated at the beginning of second clause/sentence of the relation, because otherwise the annotator should mostly recognize signals towards the end of the discourse. Our analysis evaluates and extends existing laboratory findings on DR processing by comparing a wide range of relations that are signaled not only by discourse markers. Expectation-focused discourse processing can also be explained by the ‘good-enough’ predictive approach in human language processing, which argues that humans should integrate a probabilistically ‘good-enough’ DR prediction at the boundary, and allocate more processing resource to predict the forth-coming DR (Ferreira and Lowder, 2016).

Nonetheless, this corpus study alone is not enough to prove the incrementality hypothesis in DR processing. As future work, we would also like to explore global signals, which are possibly recognized unconsciously and less likely to be identified. In addition, we plan to verify the cognitive reality of the signal positions by behavioral experiments with multiple subjects. Another goal is to design a word-level incremental discourse parser based on the findings of this work, taking into account global discourse flow.

Acknowledgments

This research was funded by the German Research Foundation (DFG) as part of SFB 1102 Information Density and Linguistic Encoding and the Cluster of Excellence (MMCI).

References

- Gerry TM Altmann. 1998. Ambiguity in sentence processing. *Trends in cognitive sciences*, 2(4):146–152.
- Lynn Carlson, Mary Ellen Okurowski, and Daniel Marcu. 2002. Ldc2002t07: Rst discourse treebank.
- Michael Collins and Brian Roark. 2004. Incremental parsing with the perceptron algorithm. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, page 111. Association for Computational Linguistics.
- Reinier Cozijn. 2000. Integration and inference in understanding causal sentences. *Unpublished doctoral dissertation, Tilburg University, Tilburg, Netherlands*.

- Dan Cristea and Bonnie Webber. 1997. Expectations in incremental discourse processing. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics and Conference of the European Chapter of the Association for Computational Linguistics*, pages 88–95. Association for Computational Linguistics.
- Debopam Das and Maite Taboada. 2017. Rst signalling corpus: a corpus of signals of coherence relations. *Language Resources and Evaluation*.
- Debopam Das, Maite Taboada, and Paul McFetridge. 2015. Ldc2015t10 rst signalling corpus.
- Vera Demberg, Frank Keller, and Alexander Koller. 2013. Incremental, predictive parsing with psycholinguistically motivated tree-adjointing grammar. *Computational Linguistics*, 39(4):1025–1066.
- F Ferreira and MW Lowder. 2016. Prediction, information structure, and good-enough language processing. *Psychology of Learning and Motivation*.
- Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, and Jun’ichi Tsujii. 2012. Incremental joint approach to word segmentation, pos tagging, and dependency parsing in chinese. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 1045–1053. Association for Computational Linguistics.
- Yangfeng Ji and Jacob Eisenstein. 2014. Representation learning for text-level discourse parsing. In *ACL (1)*, pages 13–24.
- Frank Keller. 2010. Cognitively plausible models of human language processing. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 60–67. Association for Computational Linguistics.
- Arne Köhn and Wolfgang Menzel. 2014. Incremental predictive parsing with turboparser. *Proceedings of Annual Meeting of the Association of Computational Linguistics*.
- Judith Köhne and Vera Demberg. 2013. The time-course of processing discourse connectives. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*.
- Ioannis Konstas, Frank Keller, Vera Demberg, and Mirella Lapata. 2014. Incremental semantic role labeling with tree adjointing grammar. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, volume 301.
- Qi Li and Heng Ji. 2014. Incremental joint extraction of entity mentions and relations. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.

- W Mak and T Sanders. 2010. Incremental discourse processing: How coherence relations influence the resolution of pronouns. *The linguistics enterprise: From knowledge of language to knowledge in linguistics*, pages 167–182.
- Willem M Mak and Ted JM Sanders. 2013. The role of causality in discourse processing: Effects of expectation and coherence relations. *Language and Cognitive Processes*, 28(9):1414–1437.
- William C Mann and Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281.
- Daniel Marcu. 1999. A decision-based approach to rhetorical parsing. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 365–372. Association for Computational Linguistics.
- Mitchell P Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of english: The penn treebank. *Computational linguistics*, 19(2):313–330.
- Keith K Millis and Marcel Adam Just. 1994. The influence of connectives on sentence comprehension. *Journal of Memory and Language*, 33(1):128.
- David Reitter. 2003. Rhetorical analysis with rich-feature support vector models. *Unpublished Master's thesis, University of Potsdam, Potsdam, Germany*.
- Hannah Rohde. 2008. *Coherence-Driven Effects in Sentence and Discourse Processing*. Ph.D. thesis, University of California, San Diego.
- Kenji Sagae. 2009. Analysis of discourse structure with syntactic dependencies and data-driven shift-reduce parsing. In *Proceedings of the 11th International Conference on Parsing Technologies*, pages 81–84. Association for Computational Linguistics.
- Andreas Stolcke. 1995. An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational linguistics*, 21(2):165–201.
- Michael K Tanenhaus, Michael J Spivey-Knowlton, Kathleen M Eberhard, and Julie C Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217):1632.
- Matthew J Traxler, Anthony J Sanford, Joy P Aked, and Linda M Moxey. 1997. Processing causal and diagnostic statements in discourse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(1):88.
- Junsheng Zhou, Feiyu Xu, Hans Uszkoreit, Weiguang Qu, Ran Li, and Yanhui Gu. 2016. Amr parsing with an incremental joint model. *Proceedings of SemEval*.