

The Instantiation Discourse Relation: A Corpus Analysis of Its Properties and Improved Detection

Junyi Jessy Li and Ani Nenkova

University of Pennsylvania

{ljunyi, nenkova}@seas.upenn.edu

Abstract

INSTANTIATION is a fairly common discourse relation and past work has suggested that it plays special roles in local coherence, in sentiment expression and in content selection in summarization. In this paper we provide the first systematic corpus analysis of the relation and show that relation-specific features can improve considerably the detection of the relation. We show that sentences involved in INSTANTIATION are set apart from other sentences by the use of gradable (subjective) adjectives, the occurrence of rare words and by different patterns in part-of-speech usage. Words across arguments of INSTANTIATION are connected through hypernym and meronym relations significantly more often than in other sentences and that they stand out in context by being significantly less similar to each other than other adjacent sentence pairs. These factors provide substantial predictive power that improves the identification of implicit INSTANTIATION relation by more than 5% F-measure.

1 Introduction

In an INSTANTIATION relation, one text span explains in further detail the events, reasons, behaviors and attitudes mentioned in the other (Miltsakaki et al., 2008), as illustrated by the segments below:

[a] Other fundamental “reforms” of the 1986 act have been threatened as well.

[b] The House seriously considered raising the top tax rate paid by individuals with the highest incomes.

Sentence [a] mentions “other reforms” and a threat to them, but leaves unspecified what are the reforms

or how they are threatened. Sentence [b] provides sufficient detail for the reader to infer more concretely what has happened.

The INSTANTIATION relation has some special properties. A study of discourse relations as indicators for content selection in single document summarization revealed that the first sentences from INSTANTIATION pairs are included in human summaries significantly more often than other sentences (Louis et al., 2010) and that being a first sentence in an INSTANTIATION relation is *the* most powerful indicator for content selection related to discourse relation sense. The sentences between which the relation holds also contain more sentiment expressions than other sentences (Trnavac and Taboada, 2013), making it a special target for sentiment analysis applications. Moreover, INSTANTIATION relations appear to play a special role in local coherence (Louis and Nenkova, 2010), as the flow between INSTANTIATION sentences is not explained by the major coherence theories (Kehler, 2004; Grosz et al., 1995). Many of the sentences in INSTANTIATION relation contain entity instantiations (complex examples of set-instance anaphora), such as “several EU countries”—“the UK”, “footballers”—“Wayne Rooney” and “most cosmetic purchase”—“lipstick” (McKinlay and Markert, 2011), raising further questions about the relationship between INSTANTIATIONS and key discourse phenomena.

Detecting an INSTANTIATION, however, is hard. In the Penn Discourse Treebank (PDTB) (Prasad et al., 2008), INSTANTIATION is one of the few relations that are more often *implicit*, i.e., expressed without a discourse marker such as “for exam-

ple”. Identifying implicit discourse relation is an acknowledged difficult task (Braud and Denis, 2015; Ji and Eisenstein, 2015; Rutherford and Xue, 2014; Biran and McKeown, 2013; Park and Cardie, 2012; Lin et al., 2009; Pitler et al., 2009), but the challenge is exacerbated due to the lack of *explicit* INSTANTIATIONS: explicit relations are shown to improve their implicit counterparts using data source expansion (Rutherford and Xue, 2015). Moreover, detecting INSTANTIATION also involves the skewed class distribution problem (Li and Nenkova, 2014a) because although it is one of the largest class of implicit relations, it constitutes less than 10% of all the implicit relations annotated in the PDTB.

In this work, we identify a rich set of factors that sets apart each sentence in an implicit INSTANTIATION and the pair as a whole. We show that these factors improve the identification of implicit INSTANTIATION by at least 5% in F-measure and 8% in balanced accuracy compared to prior systems.

2 Presence of INSTANTIATION

We use the Penn Discourse Tree Bank (PDTB) (Prasad et al., 2008) for the analysis and experiments presented in this paper. There are a total of 1,747 INSTANTIATION relations in the PDTB, of which 83% are implicit. INSTANTIATIONS make up 8.7% of all implicit relations and is the 5th largest among the 16 second-level relations in the PDTB.

3 Characteristics of INSTANTIATION

We identify significant factors¹ that characterize: (i) s_1 and s_2 : the first and second sentence in an INSTANTIATION pair vs. all other sentences; (ii) s_1 vs. s_2 : adjacent sentence pairs in INSTANTIATION relation vs. all other adjacent sentence pairs.

Our analysis is conducted on the PDTB except section 23, which is reserved for testing as in prior work (Lin et al., 2014; Biran and McKeown, 2015). In total, there are 1,337 INSTANTIATION sentence pairs and 43,934 non-INSTANTIATION sentences for the corpus analysis.

¹ $p < 0.05$ according to paired Wilcoxon signed rank test for real valued comparison between the two sentences in a relation, non-paired Wilcoxon rank sum for real valued factors in different types of sentences, and Kruskal-Wallis for binary valued features across different types of sentences.

	s_1	s_2	\neg Inst.
#words/sent	18.4*	26.8*	23.9

Table 1: Average # words. [*]: significant ($p < 0.05$) compared to non-instantiation sentences.

	s_1	s_2	\neg Inst.
%oov/sent	0.68*	1.54	1.46

Table 2: Average % of rare words per sentence. [*]: significant ($p < 0.05$) compared to non-instantiation sentences.

Sentence length. Intuitively, longer sentences are more likely to involve details. Table 1 demonstrates that there is an average of 8.4-word difference in length between the two sentences in an INSTANTIATION relation; moreover, s_1 s are significantly shorter (more than 5 words on average) than other sentences, and s_2 s are significantly longer.

Rare words. For each sentence, we compute the percentage of words that are not present in the 400K vocabulary of the Glove vector representations (Pennington et al., 2014). Table 2 shows that s_1 of INSTANTIATIONS contain significantly *fewer* out-of-vocabulary words compared to either s_2 and non-INSTANTIATIONS. We also compare the difference in unigram probability² of content word pairs across sentence pairs, i.e., $(w_i, w_j), w_i \in s_1, w_j \in s_2$. Compared to non-INSTANTIATION, words across INSTANTIATION arguments show significantly larger average unigram log probability difference (1.24 vs. 1.22). These numbers show that the first sentences of INSTANTIATION do not involve many unfamiliar words — an indication of higher readability (Pitler and Nenkova, 2008).

Gradable adjectives. The use of gradable adjectives (Frazier et al., 2008; de Marneffe et al., 2010)—*popular, high, likely*— may require further explanation to justify the appropriateness of their use. Here we compute the average percentage of gradable adjectives in a sentence. The list of adjectives is from Hatzivassiloglou and Wiebe (2000) and the respective percentages are shown in Table 3. Compared to other sentences, s_1 of INSTANTIATION involves significantly more gradable adjectives.

²We use a unigram language model on year 2006 of the New York Times Annotated Corpus (Sandhaus, 2008).

	s_1	s_2	\neg Inst.
%gradable adj	2.96*	2.22	2.22

Table 3: Average % of gradable adjectives per sentence. [*]: significant ($p < 0.05$) compared to non-instantiation sentences.

$s_1 > s_2$	CC EX JJR JJS NNS PDT RB [†] RBR VBG VBN VBP VBZ [†]
$s_1 < s_2$	NN NNP [†] PRP TO VBD WRB
s_1 vs \neg Inst.	CD ⁻ JJ ⁻ MD ⁻ NN ⁻ NNP ⁻ NNS ⁺ PRP ⁻ RB ⁺ TO ⁻ VB ⁻ VBD ⁻ VBG ⁺ VBP ⁺ VBZ ⁺ WDT ⁻
s_2 vs \neg Inst.	CD ⁺ DT ⁺ MD ⁻ NNP ⁺ NNS ⁺ PRP ⁺ RB ⁻ VB ⁻ VBN ⁻

Table 4: POS tags significantly different in percentage compared to non-instantiation. [†]: significance in non-instantiation pairs in the other direction. [+/-]: used more/less often than non-instantiation.

Parts of speech. We study word categories that are heavily or rarely used with INSTANTIATION by inspecting the percentage of part-of-speech tags found in each sentence. In Table 4, we show POS tags whose presence is significantly different across arguments in INSTANTIATION but not so across non-INSTANTIATION, with significance in non-INSTANTIATION in the reverse direction denoted in †. Four cases of POS occurrences are inspected:

- more often in s_1 compared to s_2 ,
- more often in s_2 compared to s_1 ,
- more (+) or less (-) in s_1 compared to non-INSTANTIATION,
- more (+) or less (-) in s_2 compared to non-INSTANTIATION.

We see that s_1 of INSTANTIATION contains more characteristic POS usage than s_2 . There are more comparative adjectives and adverbs as well as fewer nouns in s_1 compared to s_2 in INSTANTIATION pairs. The usage of verbs is also different between two arguments and s_1 contains more conjunctions and existential there. On the other hand, s_2 contains more nouns, numbers, determiners, personal pronouns and proper nouns, intuitively associated with the presence of detailed information.

Wordnet relations. Here we consider word-level relationships across arguments using Wordnet (Fellbaum, 1998). For each noun, verb, adjective and

Relation (pos)	Inst.	\neg Inst.
hypernym (n)	18.01*	21.63
meronym (n)	18.66*	15.48
holonym (n)	17.14	15.07
indirect hypernym (n)	19.17	21.03
hyponym (n)	20.33	21.75
group (v)	72.5*	68.7
indirect hypernym (v)	38.5*	41.7
hypernym (v)	76.25	74.79
hyponym (v)	80.76	78.44
entailment (v)	4.94	4.18
cause (v)	17.65	17.00
similar to (adj)	3.78	2.77
also see (adj)	6.25	5.78

Table 5: Percentage of sentence pairs with Wordnet relationships. [*]: significant ($p < 0.05$) compared to non-instantiation sentence pairs.

	s_1	s_2	Δ sim
Inst.	0.282*	0.275*	0.007
\neg Inst.	0.390	0.358	0.042*

Table 6: Average Jaccard similarity of an adjacent sentence pair s_1, s_2 with immediate context. [*]: significant ($p < 0.05$) compared to non-instantiation sentence pairs.

adverb content word pairs across arguments, we calculate the percentage of sentences with each type of Wordnet relation. Shown in Table 5, among INSTANTIATION sentence pairs there are significantly more noun-noun pairs with hypernym or meronym relationships and verbs with indirect hypernym relationship. We also observe significantly more semantically similar verbs (*group (v)*).

Lexical similarity. Finally, we inspect the similarity between sentences in each pair as well as between each sentence in a pair and their immediate prior context; specifically:

- Between s_1 and s_2 ;
- Between s_1 and C and between s_2 and C , where C denotes up to two sentences immediately before s_1 .

We compute the Jaccard similarity between sentences using their nouns, verbs, adjectives and adverbs. INSTANTIATION arguments are significantly less similar than other adjacent sentence pairs (0.335 vs. 0.505), indicating higher difference in content.

System	P	R	F	BA
Inst. specific	0.3072	0.6986	0.4268	0.7862
Vote (L&N)	0.3052	0.6438	0.4141	0.7632
L&N	0.3028	0.4521	0.3626	0.6843
B&M	0.2542	0.2055	0.2273	0.5786
Lin et al.	0.5500	0.1507	0.2366	0.5704
Brown-concat	0.1333	0.3836	0.1979	0.5919

Table 7: Precision, recall, F-measure and balanced accuracy of identifying INSTANTIATION.

Shown in Table 6, both arguments of INSTANTIATION are less similar to the immediate context. While other sentence pairs follow the pattern that s_2 is much less similar to s_1 's immediate context, this phenomenon is not significant for INSTANTIATION.

4 Experiments

In this section, we demonstrate the benefit of exploiting INSTANTIATION characteristics in the identification of the relation.

Settings. Following prior work that identifies the more detailed (second-level) relations in the PDTB (Biran and McKeown, 2015; Lin et al., 2014), we use sections 2-21 as training, section 23 as testing. The rest of the corpus is used for development. The task is to predict if an implicit INSTANTIATION relation holds between pairs of adjacent sentences in the same paragraph. Sentence pairs with INSTANTIATION relation constitute the positive class; all other non-explicit relations³ constitute the negative class. We use Logistic Regression with class weights inversely proportional to the size of each class.

Features. The factors discussed in § 3 are adopted as the *only* features in the classifier. We use the average values of s_1 and s_2 and their difference for: the number of words, difference in number of words compared to the sentence before s_1 , the percentage of OOVs, gradable adjectives, POS tags and Jaccard similarity to immediate context. We use the minimum, maximum and average differences in word-pair unigram log probability, and average Jaccard similarity across sentence pairs. For Wordnet relations, we use binary features indicating the presence of a relation.

³including AltLex, EntRel and NoRel

Results. To compare with our INSTANTIATION-specific classifier (*Inst. specific*), we show results from two state-of-the-art PDTB discourse parsers that identify second-level relations: Biran and McKeown (2015) (*B&M*) and Lin et al. (2014). We also compare the results with the classifier from our prior work (Li and Nenkova, 2014b) (*L&N*). In that work we introduce syntactic production-stick features, which minimize the occurrence of zero values in instance representation. Furthermore, we re-implemented Brown-cluster features (concatenation of clusters in each sentence) that have been shown to perform well in identifying INSTANTIATION's parent class EXPANSION (Braud and Denis, 2015).⁴

Table 7 shows the precision, recall, F-measure and balanced accuracy (average of the accuracies for the positive and negative class respectively) for each system. We show balanced accuracy rather than overall accuracy due to the highly skewed class distribution. For *Inst. specific*, we use a threshold of 0.65 for positive labels⁵. We also use it along with L&N for a *soft voting* classifier where the label is assigned to the class with larger weighted posterior probability sum from each classifier⁶. Both classifiers achieved at least 5% improvement of F-measure and 8%-10% improvement of balanced accuracy compared to other systems. These improvements mostly come from a dramatic improvement in recall. The improvement achieved by the voting classifier also indicate that *Inst. specific* provide complementary signals to syntactic production rules. Note that compared to Lin et al., *Inst. specific* behaves very differently in precision and recall, indicating potential for further system combination.

Finally, we analyze the confusion matrix induced by false positives and false negatives across Lin et al., B&M, *Inst. specific* and soft vote⁷. In Table 8, we list relations contributing at least 10% to false positives for at least one system. Remarkably, INSTANTIATION is consistently confused with

⁴The dimension of clusters are tuned on the development set. As in prior work, we use clusters in Turian et al. (2010).

⁵Tuned on development set.

⁶The weights are: 0.9 for L&N and 1.0 for *Inst. specific*, tuned on development set. We also tried voting with Brown-concat but it did not outperform combining with L&N.

⁷For other systems, we did not perform full implicit discourse parsing.

Relation	Lin et al.	B&M	Inst.	Vote
Restatement	33.3	34.1	28.7	30.8
Cause	22.2	25.0	33.9	33.6
Contrast	22.2	6.8	9.6	10.3
Conjunction	11.1	4.5	6.1	2.8
EntRel	11.1	18.2	17.4	19.6

Table 8: Relations involved in false positives, $\geq 10\%$ for least one system.

a constant set of relations: RESTATEMENT, CAUSE and EntRel. Different from other systems, the *Inst. specific* classifier demonstrates more confusion towards CAUSE than RESTATEMENT. On the false negative side, all 62 mistakes are not annotated (i.e., NoRel/EntRel) in Lin et al. For the 58 false negatives in B&M, we observe above 10%: NoRel/EntRel (56.9%), CAUSE (19.0%), RESTATEMENT (13.8%), consistent with relations relations involved in false positives.

5 Conclusion

We have characterized the implicit INSTANTIATION relation by studying significant factors that discriminate individual arguments and the sentence pairs connected by the relation. We show distinctive patterns in sentence length, word usage, semantic relationships between words as well as cross-argument and contextual similarity associated with INSTANTIATION. Using these factors as features we demonstrate significant improvement on the detection of implicit INSTANTIATION relation.

References

Or Biran and Kathleen McKeown. 2013. Aggregated word pair features for implicit discourse relation disambiguation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: Short Papers*, pages 69–73.

Or Biran and Kathleen McKeown. 2015. PDTB discourse parsing as a tagging task: The two taggers approach. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 96–104.

Chloé Braud and Pascal Denis. 2015. Comparing word representations for implicit discourse relation classification. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2201–2211.

Marie-Catherine de Marneffe, Christopher D. Manning, and Christopher Potts. 2010. “Was it good? It was provocative.” Learning the meaning of scalar adjectives. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 167–176.

Christiane Fellbaum. 1998. *WordNet*. Wiley Online Library.

Lyn Frazier, Charles Clifton Jr., and Britta Stolterfoht. 2008. Scale structure: Processing minimum standard and maximum standard scalar adjectives. *Cognition*, 106(1):299 – 324.

Barbara J. Grosz, Scott Weinstein, and Aravind K. Joshi. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2).

Vasileios Hatzivassiloglou and Janyce M. Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the 18th Conference on Computational Linguistics - Volume 1*, pages 299–305.

Yangfeng Ji and Jacob Eisenstein. 2015. One vector is not enough: Entity-augmented distributed semantics for discourse relations. *Transactions of the Association for Computational Linguistics*, 3:329–344.

Andrew Kehler. 2004. Discourse coherence. *The handbook of pragmatics*, pages 241–265.

Junyi Jessy Li and Ani Nenkova. 2014a. Addressing class imbalance for improved recognition of implicit discourse relations. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 142–150.

Junyi Jessy Li and Ani Nenkova. 2014b. Reducing sparsity improves the recognition of implicit discourse relations. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 199–207.

Ziheng Lin, Min-Yen Kan, and Hwee Tou Ng. 2009. Recognizing implicit discourse relations in the Penn Discourse Treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 343–351.

Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. 2014. A PDTB-styled end-to-end discourse parser. *Natural Language Engineering*, 20:151–184.

Annie Louis and Ani Nenkova. 2010. Creating local coherence: An empirical assessment. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 313–316.

Annie Louis, Aravind Joshi, and Ani Nenkova. 2010. Discourse indicators for content selection in summarization. In *Proceedings of the SIGDIAL 2010 Conference*, pages 147–156.

- Andrew McKinlay and Katja Markert. 2011. Modelling entity instantiations. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, pages 268–274.
- Eleni Miltsakaki, Livio Robaldo, Alan Lee, and Aravind Joshi. 2008. Sense annotation in the Penn Discourse Treebank. In *Computational Linguistics and Intelligent Text Processing*, volume 4919 of *Lecture Notes in Computer Science*, pages 275–286. Springer Berlin Heidelberg.
- Joonsuk Park and Claire 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*, pages 108–112.
- Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. *Proceedings of the Empirical Methods in Natural Language Processing*, 12:1532–1543.
- Emily Pitler and Ani Nenkova. 2008. Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 186–195.
- Emily Pitler, Annie Louis, and Ani 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*, pages 683–691.
- 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 the International Conference on Language Resources and Evaluation*.
- Attapol Rutherford and Nianwen Xue. 2014. Discovering implicit discourse relations through brown cluster pair representation and coreference patterns. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 645–654.
- Attapol Rutherford and Nianwen Xue. 2015. Improving the inference of implicit discourse relations via classifying explicit discourse connectives. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 799–808.
- Evan Sandhaus. 2008. The New York Times Annotated Corpus LDC2008T19. *Linguistic Data Consortium, Philadelphia*.
- Radoslava Trnavac and Maite Taboada. 2013. Discourse relations and affective content in the expression of sentiment in texts. In *11th ICGL Conference–Workshop on The semantic field of emotions: Interdisci.*
- Joseph Turian, Lev-Arie Ratinov, and Yoshua Bengio. 2010. Word representations: A simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 384–394.