Discourse Relations Classification and Cross-Framework Discourse Relation Classification Through the Lens of Cognitive Dimensions: An Empirical Investigation

Yingxue Fu
School of Computer Science
University of St Andrews
KY16 9SX, UK
yf30@st-andrews.ac.uk

Abstract

Existing discourse formalisms use different taxonomies of discourse relations, which require expert knowledge to understand, posing a challenge for annotation and automatic classification. We show that discourse relations can be effectively captured by some simple cognitively inspired dimensions proposed by Sanders et al. (2018). Our experiments on cross-framework discourse relation classification (PDTB & RST) demonstrate that it is possible to transfer knowledge of discourse relations for one framework to another framework by means of these dimensions, in spite of differences in discourse segmentation of the two frameworks. This manifests the effectiveness of these dimensions in characterizing discourse relations across frameworks. Ablation studies reveal that different dimensions influence different types of discourse relations. The patterns can be explained by the role of dimensions in characterizing and distinguishing different relations. We also report our experimental results on automatic prediction of these dimensions.

1 Introduction

Discourse relations are useful for various downstream NLP tasks, such as text generation (Ji and Huang, 2021) and machine translation (Sim Smith, 2017). However, discourse relations are shaped by multiple sources of information and require expert knowledge for annotation. Since the release of the Penn Discourse Treebank 2.0 (PDTB 2.0) (Prasad et al., 2008), less than 8% improvement has been made in English implicit relation classification in more than ten years (Atwell et al., 2021). Even with the development of contextualized embeddings, this task shows the least improvement in performance compared with other NLP tasks.

Another issue is that existing studies on discourse relation classification are separated into several independent strands of work (Zeldes et al., 2021). The complex nature of discourse gives rise to discourse annotation frameworks which vary in assumptions and definitions of fundamental aspects of discourse, such as what constitutes a discourse relation, what is a basic discourse unit, full-coverage or shallow discourse annotation, and how discourse structure is represented (Fu, 2022).

The leading examples of these annotation frameworks include the Rhetorical Structure Theory (RST) (Mann and Thompson, 1988), the Segmented Discourse Representation Theory (SDRT) (Asher and Lascarides, 2003) and the Discourse Lexicalized Tree-Adjoining Grammar (D-LTAG) (Forbes et al., 2003). These three frameworks have been used in various discourse annotation projects covering different languages. Based on the RST framework, the Rhetorical Structure Theory Discourse Treebank (RST-DT) (Carlson et al., 2001) is developed. SDRT forms the theoretical framework for the ANNODIS corpus (Afantenos et al., 2012), the STAC corpus (Asher et al., 2016) and so on, and D-LTAG is the theoretical foundation for PDTB (Prasad et al., 2008, 2018), which is the largest corpus annotated with discourse relations.

To enable different strands of research to come together and benefit from data across frameworks, we need an interface with which discourse relation classification tasks under different frameworks can be formulated in similar terms, independent of the underlying theoretical assumptions (Zeldes et al., 2021). The UniDim proposal by Sanders et al. (2018) represents one of the influential approaches for this task. The intuition is that discourse relations of different frameworks can be decomposed into cognitive primitives rooted in the Cognitive approach to Coherence Relations (CCR) (Sanders et al., 1992, 1993) (hence denoted as the CCR framework), and people can make use of these elementary notions to relate and compare discourse relations. These primitives are not intended to form a complete and descriptively adequate account of discourse relations but are targeted at a psychologically plausible theory of discourse relations (Sanders et al., 1992). Additional primitives are added in later studies to reach better linguistic and cognitive coverage (Crible and Degand, 2019).

Sanders et al. (2018) and other researchers such as Rehbein et al. (2016) try to test if discourse relations annotated based on the CCR framework are consistently categorized into relations under other frameworks. Their investigation reveals that discrepancies between frameworks arise due to variations in how coherence relations are defined, the methods used to perform the annotation, and the rules governing segmentation, and the alignment of discourse relations is generally many-to-many.

In this study, we aim to assess to what extent these CCR dimensions provide information about discourse relations of different frameworks. We assume that CCR dimensions are annotated in parallel to discourse relation annotations of other frameworks and utilize these dimensions as features in discourse relation classification tasks. The improvement/degradation of performance relative to the case without such features as a measure of the information that these dimensions provide. In this way, we show empirical evidence of the effectiveness of the UniDim proposal in representing and bridging discourse relations of different discourse annotation schemes.

Our contributions include:

- We show that the dimensions of the UniDim proposal effectively capture discourse relations and are useful for training computational systems for discourse relation classification, both for RST relation classification and PDTB explicit and implicit relation classification, yielding significant performance gains. Such elementary cognitive dimensions can be useful features for the challenging task of discourse relation classification.

- We demonstrate that these dimensions can work as an interface for discourse relations across different frameworks. It is possible to train one discourse relation classification model on PDTB and apply the model to the discourse relation classification task in RST with transfer learning and the performance is as high as training a model specifically for RST relation classification, in spite of differences in discourse segmentation between the two frameworks. The CCR dimensions provide an effective means of bridging discourse relations of different frameworks.

- We report experimental results on automatic prediction of these dimensions with RST-DT, PDTB 3.0 and a combination of the two corpora.

2 Related Work

2.1 Mapping Discourse Relations of Different Frameworks

Prior studies on mapping discourse relations of different frameworks adopt varied approaches. Some researchers propose common inventories of relations that are created based on analysis of discourse relations of different frameworks (Benamara and Taboada, 2015; Bunt and Prasad, 2016). Alternatively, an intermediate representation may be used to reduce the number of mappings necessary to harmonize different frameworks (Chiarcos, 2014; Sanders et al., 2018). As there are corpora that contain parallel annotations under different frameworks on the same texts, these corpora are used to identify mappings between discourse relations. Since this approach relies on textual matching, differences in discourse segmentation would hinder relation mapping, leaving only a small number of relations successfully mapped between different frameworks (Bourgonje and Zolotarenko, 2019; Scheffler and Stede, 2016). The study by Demberg et al. (2019) employs the *strong nuclearity hypothesis* (Marcu, 2000) to mitigate this problem. Demberg et al. (2019) show that the Unified Dimension (UniDim) approach is relatively successful in mapping relations between RST-DT and PDTB 2.0.

Roze et al. (2019) investigates the possibility of predicting CCR dimensions automatically. They achieve an accuracy above the baseline of majority class guessing. Furthermore, they try to predict relations of PDTB 2.0 from these dimensions, and it is shown that the accuracy is much lower than that of training a model for predicting PDTB relations directly. The low performance may be attributed to the high level of under-specification in the mapping from PDTB relations to these dimensions and the reverse mapping from dimension combinations to the hierarchical PDTB sense labels, especially when the mapping is not necessarily one-to-one.

Recent studies propose to represent discourse relations as question-answering (QA) pairs (Ko et al., 2022; Pyatkin et al., 2020). While this approach is designed to simplify discourse relation labelling, some relations cannot be expressed by QA pairs (Pyatkin et al., 2020), and evaluation is difficult. Moreover, open-ended QA leads to annotations similar to the GraphBank (Wolf and Gibson, 2004), which has higher complexity than the other frameworks.

2.2 Dimensions in UniDim Proposal

The main approach adopted in the UniDim proposal is to use cognitively inspired dimensions as an intermediate representation and decompose discourse relations of different frameworks into these dimensions so that they can be related and compared. The result contains five dimensions which are rooted in the Cognitive approach to Coherence Relations (CCR) (Sanders et al., 1992, 1993) and some additional dimensions that are added to allow more relations to be better represented (collectively referred to as "UniDim dimensions" or "dimensions of the UniDim proposal" in the following). We give an overview of these dimensions here.

Two segments that may stand in a discourse relation are identified first, the two segments being denoted as S_1 and S_2 in linear order, and the underlying propositions being denoted as P and Q in linear order.

The first dimension is **basic operation**, which has two values: *causal* and *additive*. A causal relation means that the two segments are strongly connected and typically, an implication relation $P \rightarrow Q$ can be deduced. In (1), S_2 shows the cause and S_1 gives the consequent. If the two segments are just loosely connected and only a conjunction relation $P \wedge Q$ can be inferred, the value at this dimension is additive, as shown in (2).

(1) [He immigrated to the US,]$_{S_1}$ because [his natural parents were believed to live there.]$_{S_2}$

(2) [She is a painter]$_{S_1}$ and [her studio is a few blocks away.]$_{S_2}$

As indicated in Sanders et al. (2018), basic operation can be used to distinguish causal relations or conditional relations from additive relations or temporal relations.

The second dimension is **source of coherence**. It has two values: semantic and pragmatic in the original proposal (Sanders et al., 1992), later renamed as *objective* and *subjective* in Maat and Sanders (2000), respectively. A relation is objective if the segments are connected because of their propositional content, and the relation holds because the connection is coherent based on world knowledge, as shown in (3). A relation is subjective if the speaker's reasoning or the pragmatic effect of the relation is prominent. (4) shows a claim in S_2 and S_1 is an argument that supports it.

(3) [It was dark outside,]$_{S_1}$ so [he lit up a candle.]$_{S_2}$

(4) [Smoking is unhealthy]$_{S_1}$ and [we should put a limit on it.]$_{S_2}$

This dimension can be used to distinguish relations that are related to real-world situations, such as temporal sequence, and cause-consequence, from argumentative relations, such as claim-argument or evidence-justification (Sanders et al., 2018).

The third dimension is **implication order**. This dimension distinguishes between *non-basic* and *basic* orders of causal relations, and does not apply to additive relations, which are generally symmetric. For a causal relation characterized by $P \rightarrow Q$, if S_1 expresses P and S_2 expresses Q (note that S_1 and S_2 are in linear order), then this relation is in basic order, as shown in (6). If S_2 actually expresses P while S_1 expresses Q, this relation is in non-basic order, as shown in (5).

(5) [He did not attend the conference,]$_{S_1}$ because [he received a message telling him not to go.]$_{S_2}$

(6) Because [he received a warning message,]$_{S_1}$ [he did not attend the conference.]$_{S_2}$

It is clear to see that the implication order dimension is mainly used to distinguish relations with directionality, such as cause-result and cause-reason.

The fourth dimension is **polarity**. A relation is characterized by *positive* polarity if the propositions P and Q, expressed by S_1 and S_2, respectively, have the same logical polarity and support each other, as shown in (7). A relation is of *negative* polarity if the relation involves the juxtaposition of $\neg P$ and P or $\neg Q$ and Q in the two segments, as shown in (8). In this example, a positive polarity would require a reason or result that supports the decision of closing the library.

(7) [We like the garden]$_{S_1}$ because [it is pretty.]$_{S_2}$

(8) [The university library was closed]$_{S_1}$ although [students wanted more space for study.]$_{S_2}$

This dimension is useful for capturing contrastive, adversative and concession relations (Sanders et al., 2018).

The fifth dimension is **temporality**, which distinguishes between *temporal* and *non-temporal* relations. Under temporal relations, temporality has three values: *synchronous*, *chronological* and *anti-chronological*. Synchronous relations are those temporal relations which feature simultaneous occurrence of events. If events described in the segments happen in temporal order, then the relation is chronological, otherwise the relation is anti-chronological.

In order to characterize more relations, additional dimensions are introduced, including **specificity**, **lists** and **alternatives** for additive relations, and **conditionals** and **goal-oriented relations** for causal relations (denoted collectively as "additional dimensions" in the following).

3 Methodology

Since RST-DT and PDTB both use the WSJ articles of the Penn Treebank, cross-framework relation classification of RST and PDTB by automatic means would be less influenced by domain shift. Therefore, we focus on the two frameworks. For PDTB, we use PDTB 3.0, which is newer and introduced systematic changes.

As we are primarily interested in the effectiveness of UniDim dimensions rather than improving algorithms for discourse relation classification, simple models are implemented in the experiments.

3.1 Discourse Relation Classification

Discourse relation classification is a typical multi-class classification task. Given a span/argument pair with tokens $S = [CLS]$, $S_1^{(1)}$ ... $S_m^{(1)}$, $[SEP]$, $S_1^{(2)}$ ... $S_n^{(2)}$, we obtain the representation of the sequence from a pre-trained language model, denoted as $f_{PLM}(S)$, and the embeddings of the dimensions E are obtained from embedding layers, where the embeddings are initialized from uniform distributions and trainable. The representation of the input and the embeddings of dimensions are concatenated:

$$h_S = f_{PLM}(S) \oplus Edim_{pol} \oplus Edim_{bop} \oplus ... \quad (1)$$

The dim_{pol} and dim_{bop} ... represents the UniDim dimensions, including polarity, basic operation, implication order, source of coherence, temporaltiy, specificity, alternative, conditional and goal.

The representation is fed to two two-layer feed-forward networks (FFNs) with LeakyReLU as activation functions:

$$\hat{h} = g_2(W_2 * g_1(W_1 * h_S)) \quad (2)$$

where g_1 and g_2 represent the non-linear activation functions of first and the second FFNs, respectively. W_1 and W_2 denote weights of the first layers of the two FFNs, and bias terms are omitted for clarity.

A classifier layer is configured on top of the second FFN. The predicted result $\hat{y}$ is obtained with:

$$\hat{y} = softmax(W_3 * \hat{h}) \quad (3)$$

Cross-entropy loss is used in the loss function:

$$\mathcal{L}_c = -\sum_{i=1}^{N}\sum_{l=1}^{C} c_l^i \log p(c_l^i) \quad (4)$$

where N is the batch size, C is the total number of classes, and $p(c_l^i)$ is the probability predicted for a class c.

In this design, we take our experiments with transfer learning for cross-framework discourse relation classification into consideration, as we try to keep the architecture and only replace the last classifier layer to fit the model on new data. Moreover, our preliminary experiments indicate that removing the second FFN causes a significant performance drop.

Baseline model The BertForSequenceClassification model from the Transformers library (Wolf et al., 2020) is used as the baseline model, in which a classifier layer is added on top of the contextualized embeddings of the input sequence. For an input sequence S, its representation is obtained with:

$$h_S = f_{PLM}(S) \quad (5)$$

The predicted result $\hat{y}$ is obtained with:

$$\hat{y} = softmax(W_b * h_S) \quad (6)$$

As shown in Kim et al. (2020), this model is a strong baseline. We use the *bert-base-uncased* BERT model in all our experiments for comparison of experimental results.

3.2 Cross-framework Discourse Relation Classification

We hypothesize that if UniDim dimensions form an effective "interlingua" of discourse relations from different frameworks, we can train a model for discourse relation classification in one framework and apply the model for relation classification in another framework without much modification. The transfer learning framework can be used for this experiment.

As PDTB 3.0 is much larger than RST-DT, a natural choice would be to treat PDTB relation classification as the source task and RST relation classification as the target task (Wang et al., 2019).

We first train a model as described in section 3.1 on all the PDTB data, and freeze all the layers but the last classifier layer so that the model can be fit on RST data.

Formally, for a pair of PDTB arguments $P = [CLS], A_1^{(1)} ... A_m^{(1)}, [SEP], A_1^{(2)} ... A_n^{(2)}$, we obtain the representation of sequence P with equation (1). Through training, the parameters in equation (2) are learnt for the PDTB relation classification task. With these parameters, for an RST span pair $R = [CLS], R_1^{(1)}, ..., R_m^{(1)}, [SEP], R_1^{(2)}, ..., R_n^{(2)}$, we first obtain the representation of sequence R with equation (1), denoted as h_R, and with the parameters learnt for PDTB relation classification, we obtain the representation $\widehat{h_R}$:

$$\widehat{h_R} = g_2(W_2 * g_1(W_1 * h_R)) \tag{7}$$

The predicted result $\hat{y}$ for RST relation classification is obtained with:

$$\hat{y} = softmax(W_r * \widehat{h_R}) \tag{8}$$

where W_r is the weight to be learnt for RST relation classification.

Baseline model As we transfer knowledge from PDTB relation classification to RST relation classification, the baseline model is a model trained specifically for RST relation classification with BERT embeddings and UniDim dimensions as input. For the baseline model in section 3.1, where only BERT embeddings are used, we train a model for PDTB relation classification and apply the model to RST relation classification without using UniDim dimensions.

3.3 Automatic UniDim Dimension Prediction

Since the dimensions may be related to each other, we train one model for predicting the nine dimensions in equation 1 together.

For an input sequence S, we obtain its representation h_S with equation 5. A two-layer FFN f with LeakyReLU activation function is applied to h_S before nine classification layers $c_{i|i=1...9}$ are applied:

$$\hat{y} = softmax(W_{c_i} * f(h_S)) \tag{9}$$

We train the model on PDTB, RST and the combination of PDTB and RST data, respectively. The results reported in Roze et al. (2019) are our baseline.

4 Experiments

We use the mapping table given in Sanders et al. (2018) (Appendix A) for obtaining the dimension values for relation labels of RST-DT. As no mapping table is provided for PDTB 3.0, we create the mapping table by ourselves (Appendix B).

4.1 Data Preprocessing

We binarize the RST trees based on the procedure in Ji and Eisenstein (2014) and extract pairs of spans that are connected by a relation. Following Sanders et al. (2018), we exclude *Same-Unit* and *Attribution* relations from RST-DT, leaving 16 relations. We use the standard split of the corpus and take 20% from the training set for validation.

Since PDTB level-2 relations carry specific and generally more useful information, we focus on level-2 relation classification for PDTB. We exclude relations that have fewer than 100 instances to alleviate data imbalance, as suggested in Kim et al. (2020). We follow the data split in Ji and Eisenstein (2015), using sections 2-20 for training, 0-1 for validation and 21-22 for testing.

We use the pre-trained BERT model (Devlin et al., 2019) for obtaining contextualized embeddings and the *[CLS]* and *[SEP]* tokens are inserted following the settings of the BERT model, which is shown to benefit inter-sentential (Shi and Demberg, 2019) and intra-sentential (Zhao and Webber, 2021) implicit discourse relation classification.

Among the UniDim dimensions, we exclude *list* because this dimension is proposed for representing the *List* relation in PDTB, which has been removed from the sense hierarchy in PDTB 3.0. Following Roze et al. (2019), we merge *specificity-example* and *specificity-equivalence* into specificity, and add the *NS* label in cases of ambiguity or under-specification. The *N.A.* label is kept when it appears on its own to reflect the fact that some dimensions do not apply to certain types of relations. The default values of additional dimensions are set to negative because they are only applicable to some relations and typically have binary values.

On the whole, the dimensions are heavily imbalanced and have high degree of under-specification. Statistics for the distribution of these dimensions are shown in Appendix C. Hyper-parameter settings and model training details are described in

Appendix D.

4.2 Evaluation

For RST relation classification, the settings of the DISRPT 2021 shared task on relation classification (Zeldes et al., 2021) are the closest to ours. We report their best accuracy on RST-DT (Gessler et al., 2021) alongside our baseline model results for comparison.

After preprocessing, we perform 12-way explicit relation classification and 14-way implicit relation classification for PDTB. While most of the previous studies use PDTB 2.0 and recent studies on PDTB 3.0 only focus on implicit relation classification, when settings of previous studies are close to ours, we report their results alongside our baseline results[1].

4.3 Results and Discussion

We report our experimental results on the test sets, which are computed with the Scikit-Learn library (Pedregosa et al., 2011). We can expect that RST and PDTB data show different patterns. For RST, the dimension values for end labels may be clear, but when end labels are grouped into a class, the values could be rather mixed. For PDTB, as L2 sense classification is performed, the process of grouping relations into broader classes happens at L3, which only encodes directionality, and dimensions that are related to directionality are affected, such as implication order, but the other dimensions are not influenced. Therefore, dimension values for PDTB classes tend to be less ambiguous. Moreover, data amount differences are likely to have notable influence on the results. We do not report the results of additional dimensions separately because their individual effects are not obvious.

4.3.1 RST Relation Classification

Table 1 shows results on RST-DT. When UniDim dimensions are added as features, a significant performance gain can be obtained. Some relations can be recognized with 100% accuracy. However, relations including *Comparison*, *Manner-means*, *Summary* and *Textual-Organization* cannot be recognized. From Fig. 3 in Appendix E, it is clear that these relations have small amounts of training data. As we focus on broader classes rather than end labels in relation classification, we can see from the mapping table in Appendix A that dimension values under these classes are mixed. It is difficult for the model to learn patterns from the data.

[1] We build and run all the baseline models mentioned in section 3.1 and section 3.2 by ourselves.

	P	R	$F1$	$P_{b.}$	$R_{b.}$	$F1_{b.}$	$C.$
Background	1.00	1.00	1.00	0.47	0.35	0.40	111
Cause	0.92	0.70	0.79	0.50	0.17	0.25	82
Comparison	0.00	0.00	0.00	0.61	0.38	0.47	29
Condition	1.00	1.00	1.00	0.79	0.71	0.75	48
Contrast	0.99	1.00	0.99	0.75	0.68	0.72	146
Elaboration	0.75	1.00	0.86	0.65	0.88	0.75	796
Enablement	0.92	1.00	0.96	0.61	0.85	0.71	46
Evaluation	0.99	1.00	0.99	0.29	0.14	0.19	80
Explanation	0.72	0.97	0.83	0.46	0.27	0.34	110
Joint	1.00	0.03	0.06	0.67	0.62	0.64	212
Manner-Means	0.00	0.00	0.00	0.68	0.48	0.57	27
Summary	0.00	0.00	0.00	0.88	0.47	0.61	32
Temporal	1.00	1.00	1.00	0.74	0.27	0.40	73
Textual-Organization	0.00	0.00	0.00	0.44	0.44	0.44	9
Topic-Change	0.28	1.00	0.44	0.28	0.38	0.32	13
Topic-Comment	0.71	0.21	0.32	0.00	0.00	0.00	24
Acc.		**0.81**		0.63 (vs DISRPT 2021: 0.67)			
Macro-F1	**0.64**	**0.62**	**0.58**	0.55	0.44	0.47	1838

Table 1: Results of RST relation classification. The columns in blue show the results of our method and uncolored columns show the results of the baseline model, and the last column shows the count of occurrences of each relation in the test set. We use this convention in reporting the results.

To have a better understanding of the influence of each dimension on the results, we performed ablation studies and the results are shown in Table 2.

	Acc	P	R	$F1$
Total	0.81	0.64	0.62	0.58
-Pol.	0.74	0.49	0.48	**0.48**
-Basic Op.	0.78	0.52	0.58	0.53
-SoC.	0.78	0.52	0.58	0.53
-Impl. order	0.81	0.58	0.60	0.55
-Temp.	0.80	0.59	0.60	0.55
-Add.	0.80	0.52	0.59	0.54

Table 2: Results of ablation studies for RST relation classification, showing the overall accuracy (Acc), precision (P), recall(R) and macro-averaged F1 ($F1$) for dimensions of polarity (Pol.), basic operation (Basic Op.), source of coherence (SoC.), implication order (Impl. order), temporality (Temp.) and additional dimensions (Add.), respectively.

As shown in Table 2, removing the polarity dimension causes the biggest performance drop in macro-averaged F1. By examining the detailed results (Table 33, Appendix L), we find that removing this dimension has noticeable influence on the recognition of *Contrast*($\downarrow$ 0.41), *Evaluation*($\downarrow$ 0.26), *Topic-Change*($\downarrow$ 0.44) and *Topic-Comment*($\downarrow$ 0.32). The correlation between *Contrast* and this dimension is self-evident. Examination of the mapping table suggests that the rest of these relations have ambiguous or mixed values in the other dimensions and their data amounts are small, making it difficult for the model to learn any patterns.

4.3.2 PDTB Explicit Relation Classification

Table 3 shows the results of 12-way explicit relation classification. The overall accuracy score is high and the majority of the relations can be recognized with near perfect performance, which means that the UniDim dimensions are effective in characterizing most of the PDTB explicit relations. However, in spite of the noticeable improvement in overall accuracy, our method does not show improvement over the baseline model in macro-averaged F1 score. This is likely due to the strong reliance of pre-trained language models on lexical cues in discourse relation classification tasks (Kim et al., 2020) and these lexical cues are effective features for this task. Moreover, with our approach, the *Level-of-detail* and *Substitution* relations cannot be recognized. The two relations have the smallest data amount, and in terms of dimension values, *Substitution* is similar to *Concession* and *Level-of-detail* is similar to *Manner*. It is possible that the model predicts *Manner* for instances of *Level-of-detail*, which explains the lower precision for *Manner*.

	P	R	$F1$	$P_{b.}$	$R_{b.}$	$F1_{b.}$	$C.$
Asynchronous	1.00	1.00	1.00	0.97	0.87	0.92	127
Cause	1.00	1.00	1.00	0.82	0.89	0.85	115
Concession	0.96	1.00	0.98	0.89	0.95	0.92	285
Condition	1.00	1.00	1.00	0.93	0.92	0.93	61
Conjunction	1.00	1.00	1.00	0.97	0.96	0.96	516
Contrast	1.00	1.00	1.00	0.52	0.48	0.50	50
Disjunction	1.00	1.00	1.00	0.90	1.00	0.95	18
Level-of-detail	0.00	0.00	0.00	0.71	0.75	0.73	20
Manner	0.35	1.00	0.52	0.42	0.91	0.57	11
Purpose	1.00	1.00	1.00	0.62	0.45	0.52	29
Substitution	0.00	0.00	0.00	1.00	0.92	0.96	13
Synchronous	1.00	1.00	1.00	0.81	0.71	0.76	126
Acc.		**0.98**			0.89		
Macro-F1	**0.78**	**0.83**	**0.79**	0.80	0.82	0.80	1371

Table 3: Results of PDTB explicit relation classification.

The results of ablation studies are shown in Table 4. Removing the source of coherence dimension causes the biggest performance drop in macro-averaged F1. Through examining the detailed results, we find that without this dimension, the *Disjunction* relation cannot be recognized. Meanwhile, removing this dimension causes a drop of 0.15 for identifying the *Contrast* relation and a drop of 0.14 for recognizing the *Synchronous* relation. The *Disjunction* relation has a small data amount, and the model might predict *Contrast* for instances of *Disjunction*, since they are similar in the absence of this dimension, which may account for the lower precision for *Contrast*.

	Acc	P	R	$F1$
Total	0.98	0.78	0.83	0.79
-Pol.	0.95	0.74	0.81	0.76
-Basic Op.	0.98	0.78	0.83	0.79
-SoC.	0.94	0.67	0.73	**0.68**
-Impl. order	0.98	0.78	0.83	0.79
-Temp.	0.95	0.76	0.81	0.77
-Add.	0.96	0.73	0.73	0.73

Table 4: Results of ablation studies for PDTB explicit relation classification.

4.3.3 PDTB Implicit Relation Classification

Table 5 shows the results of 14-way implicit relation classification. The previous best result under similar settings is 0.64 in overall accuracy (Kim et al., 2020), which is achieved with *large-cased* XLNet (Yang et al., 2019). Our baseline 56% accuracy is consistent with the results in Kim et al. (2020).

	P	R	$F1$	$P_{b.}$	$R_{b.}$	$F1_{b.}$	$C.$
Asynchronous	1.00	1.00	1.00	0.62	0.61	0.62	95
Cause	1.00	1.00	1.00	0.60	0.63	0.61	366
Cause+Belief	1.00	0.42	0.59	0.00	0.00	0.00	12
Concession	1.00	0.92	0.96	0.44	0.40	0.42	84
Condition	1.00	1.00	1.00	0.71	0.42	0.53	12
Conjunction	0.90	1.00	0.95	0.49	0.61	0.54	221
Contrast	0.98	1.00	0.99	0.45	0.42	0.43	50
Equivalence	0.00	0.00	0.00	0.12	0.04	0.06	24
Instantiation	0.00	0.00	0.00	0.77	0.54	0.64	107
Level-of-detail	0.60	1.00	0.75	0.45	0.48	0.46	180
Manner	0.00	0.00	0.00	0.38	0.60	0.46	15
Purpose	0.92	0.94	0.93	0.92	0.98	0.95	88
Substitution	0.75	1.00	0.86	0.43	0.48	0.45	21
Synchronous	0.87	0.97	0.92	0.27	0.10	0.15	40
Acc.		**0.87**			0.56		
Macro-F1	**0.72**	**0.73**	**0.71**	0.48	0.45	0.45	1315

Table 5: Results of PDTB implicit relation classification.

As is shown in Table 5, adding UniDim dimensions brings significant performance gain for this task, which is challenging for the baseline model. Meanwhile, we notice that relations including *Equivalence*, *Instantiation* and *Manner* are difficult to recognize. In terms of dimension values, *Equivalence* is similar to *Conjunction*, which has a much larger amount of data. It is likely that the model predicts *Conjunction* for *Equivalence*, hence the lower precision for *Conjunction*. *Instantiation*, *Manner* and *Level-of-detail* have the same dimension values, and as the data amount for *Level-of-detail* is much larger, the model may predict *Level-of-detail* for instances of the other two relations, causing the precision score for *Level-of-detail* to go down.

The results of ablation studies are shown in Table 6. Both the implication order dimension and the additional dimensions have substantial influence on the F1 score. Removing the implication order dimension does not cause much decrease in the overall accuracy score but mainly lowers the F1 score, while removing the additional dimensions reduces both the overall accuracy score and the F1 score.

	Acc	P	R	$F1$
Total	0.87	0.72	0.73	0.71
-Pol.	0.87	0.71	0.71	0.70
-Basic Op.	0.87	0.72	0.73	0.71
-SoC.	0.87	0.72	0.73	0.71
-Impl. order	0.86	0.57	0.64	**0.60**
-Temp.	0.87	0.72	0.73	0.71
-Add.	0.73	0.64	0.64	**0.62**

Table 6: Results of ablation studies for PDTB implicit relation classification.

Detailed results (Table 26 in Appendix J) show that removing the implication order dimension causes a drop of 0.07 in recognizing *Concession*, a drop of 0.86 in recognizing *Substitution* and a drop of 0.59 in recognizing *Cause+Belief*. As the last two relations cannot be recognized, the macro-averaged F1 shows a significant decrease. Similarly, this is associated with differences in data amount and how different relations can be distinguished from each other without the dimension, for instance, *Substitution* has a small data amount, and without the implication order dimension, the model might confuse this relation with *Concession* and predict *Concession* for instances of both relations, which may explain the lower precision for *Concession*. If the additional dimensions are removed, major relations that are impacted include *Condition*($\downarrow$ 0.14), *Conjunction*($\downarrow$ 0.37), and *Level-of-detail*($\downarrow$ 0.75). In this case, the *Level-of-detail* relation cannot be identified. Without this dimension, *Level-of-detail* has the same dimension values as *Conjunction*, which has a larger data amount. The model may predict *Conjunction* for both classes, which causes precision for *Conjunction* to decrease.

4.3.4 Cross-Framework Discourse Relation Classification

As RST does not distinguish explicit and implicit relations, we train a model on the whole PDTB data for the source task. We show the overall performance of transfer learning from PDTB to RST in Table 7. The settings of the DISRPT 2021 shared task are the closest to our experiments, and their best results (Gessler et al., 2021) are shown alongside the baseline model for comparison. As is clear from the table, the results of transfer learning based on the baseline BERT model show noticeable effect of negative transfer (0.63 $\rightarrow$ 0.58 in overall accuracy and 0.47 $\rightarrow$ 0.33 in F1 score), while with our method, the overall accuracy does not show any decrease and the F1 score is only 1% lower. This shows that the UniDim dimensions may serve as an effective interface for relations of different frameworks. The detailed results for the source and target tasks are shown in Tables 39 and 40 in Appendix M.

Task	Acc.	Macro-F1
target RST (BERT+Dim)	**0.81**	**0.57**
RST-specific (BERT+Dim) from Table 1	0.81	0.58
src PDTB total (BERT+Dim)	0.86	0.67
target RST (BERT only)	0.58	0.33
RST-specific (BERT only) from Table 1	0.63	0.47
src PDTB total (BERT only)	0.71 (vs. DISRPT 2021: 0.74)	0.61

Table 7: Results of transfer learning from PDTB to RST.

4.3.5 Automatic Dimension Prediction

We show our experimental results of automatic prediction of UniDim dimensions in Table 8. As is clear from the table, reasonable performance for this task can be achieved. Note that the baseline results are based on PDTB 2.0 and separate classifiers are trained for each dimension.

The performance on PDTB is higher than on RST data with the exception of *Temporality* and *Goal*. As PDTB allows multi-sense annotation, instances labeled with temporal relations might be annotated with labels of causal relations, and instances for which a *Purpose* relation can be inferred (captured by the *Goal* dimension), a *Manner* relation is also possible (not involving the *Goal* dimension), which poses a challenge for machine learning systems.

Moreover, combining the two corpora to augment training data does not improve the performance over using PDTB data alone but it is helpful for improving performance on RST data. RST data amount is much smaller and adding more data is beneficial. As relations of the two frameworks may not be completely compatible and combining the two corpora might introduce inconsistent and redundant data, combining the datasets is likely to be more useful in low-resource settings.

	PDTB		RST		PDTB+RST		Baseline	
	Acc.	Macro-F1	Acc.	Macro-F1	Acc.	Macro-F1	Acc.	Macro-F1
Pol.	0.92	0.57	0.85	0.58	0.89	0.56	0.82	0.50
Basic Op.	0.80	0.52	0.76	0.45	0.77	0.50	0.76	0.38
SoC.	0.75	0.72	0.67	0.45	0.70	0.59	0.68	0.50
Impl. order	0.76	0.50	0.75	0.38	0.75	0.48	0.78	0.41
Temp.	0.79	0.59	0.86	0.30	0.82	0.43	0.73	0.48
Spec.	0.87	0.65	0.80	0.72	0.83	0.66	0.85	-
Alter.	1.00	0.95	1.00	0.50	1.00	0.95	0.99	-
Cond.	0.99	0.86	0.98	0.83	0.98	0.83	0.99	-
Goal	0.91	0.75	0.97	0.75	0.93	0.74	-	-

Table 8: Results of UniDim dimension prediction. Blue columns show classification accuracy and grey columns show macro-averaged F1.

5 Conclusion and Future Work

By incorporating the UniDim dimensions proposed in Sanders et al. (2018) in discourse relation classification tasks, we obtain quantitative results of the effectiveness of these dimensions in capturing discourse relations of different frameworks and bridging discourse relations across frameworks. Ablation studies reveal the influence of these dimensions on different types of discourse relations. Meanwhile, we show that these dimensions can be predicted automatically with a simple model. These dimensions are potentially useful features for discourse relation classification across frameworks. Therefore, in future work, we plan to incorporate automatically predicted dimensions in our models.

6 Limitations

Since we need to create the mapping table for PDTB 3.0 by ourselves, it is unavoidable that there may be errors and inconsistencies with existing mapping tables for the other frameworks.

Meanwhile, in the mapping table provided in Sanders et al. (2018), to obtain the values of the dimensions, we need all the information of a relation label, for instance, to represent an RST relation label with dimensions, we need the nuclearity label and whether the relation is mono-nuclear or multi-nuclear in addition to the relation label itself, and in the case of a PDTB relation, we need the relation label and the order of the arguments. This is because these dimensions are not incorporated in the annotation process of RST-DT and PDTB, and only a general mapping is possible. We consider the resultant ambiguity and under-specification unavoidable.

7 Ethics Statement

This study does not involve special ethical considerations. The potential impact may include providing computational evidence of the validity of cognitive study of discourse relations and attracting attention to cognitive frameworks of discourse, which may spur fine-grained research on the correlation between cognitive dimensions and different discourse relations and how different language models perform from this perspective.

References

Stergos Afantenos, Nicholas Asher, Farah Benamara, Myriam Bras, Cécile Fabre, Mai Ho-dac, Anne Le Draoulec, Philippe Muller, Marie-Paule Péry-Woodley, Laurent Prévot, Josette Rebeyrolles, Ludovic Tanguy, Marianne Vergez-Couret, and Laure Vieu. 2012. An empirical resource for discovering cognitive principles of discourse organisation: the ANNODIS corpus. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2727–2734, Istanbul, Turkey. European Language Resources Association (ELRA).

Nicholas Asher, Julie Hunter, Mathieu Morey, Benamara Farah, and Stergos Afantenos. 2016. Discourse structure and dialogue acts in multiparty dialogue: the STAC corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2721–2727, Portorož, Slovenia. European Language Resources Association (ELRA).

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press, United States.

Katherine Atwell, Junyi Jessy Li, and Malihe Alikhani. 2021. Where are we in discourse relation recognition? In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 314–325, Singapore and Online. Association for Computational Linguistics.

Farah Benamara and Maite Taboada. 2015. Mapping different rhetorical relation annotations: A proposal. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 147–152, Denver, Colorado. Association for Computational Linguistics.

Peter Bourgonje and Olha Zolotarenko. 2019. Toward cross-theory discourse relation annotation. In *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019*, pages 7–11, Minneapolis, MN. Association for Computational Linguistics.

Harry Bunt and Rashmi Prasad. 2016. ISO DR-Core (ISO 24617-8): Core concepts for the annotation of discourse relations. In *Proceedings 12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-12)*, pages 45–54.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurovsky. 2001. Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*.

Christian Chiarcos. 2014. Towards interoperable discourse annotation. discourse features in the ontologies of linguistic annotation. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4569–4577, Reykjavik, Iceland. European Language Resources Association (ELRA).

Ludivine Crible and Liesbeth Degand. 2019. Domains and functions: A two-dimensional account of discourse markers. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics*, (24).

Vera Demberg, Merel CJ Scholman, and Fatemeh Torabi Asr. 2019. How compatible are our discourse annotation frameworks? insights from mapping rst-dt and pdtb annotations. *Dialogue & Discourse*, 10(1):87–135.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Katherine Forbes, Eleni Miltsakaki, Rashmi Prasad, Anoop Sarkar, Aravind Joshi, and Bonnie Webber. 2003. D-LTAG system: Discourse parsing with a lexicalized tree-adjoining grammar. *Journal of Logic, Language and Information*, 12(3):261–279.

Yingxue Fu. 2022. Towards unification of discourse annotation frameworks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 132–142, Dublin, Ireland. Association for Computational Linguistics.

Luke Gessler, Shabnam Behzad, Yang Janet Liu, Siyao Peng, Yilun Zhu, and Amir Zeldes. 2021. DisCoDisCo at the DISRPT2021 shared task: A system for discourse segmentation, classification, and connective detection. In *Proceedings of the 2nd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2021)*, pages 51–62, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Haozhe Ji and Minlie Huang. 2021. DiscoDVT: Generating long text with discourse-aware discrete variational transformer. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4208–4224, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Yangfeng Ji and Jacob Eisenstein. 2014. Representation learning for text-level discourse parsing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13–24, Baltimore, Maryland. Association for Computational Linguistics.

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.

Najoung Kim, Song Feng, Chulaka Gunasekara, and Luis Lastras. 2020. Implicit discourse relation classification: We need to talk about evaluation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5404–5414, Online. Association for Computational Linguistics.

Wei-Jen Ko, Cutter Dalton, Mark Simmons, Eliza Fisher, Greg Durrett, and Junyi Jessy Li. 2022. Discourse comprehension: A question answering framework to represent sentence connections. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 11752–11764, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Henk Pander Maat and Ted Sanders. 2000. Domains of use or subjectivity? the distribution of three dutch causal connectives explained. *Topics in English Linguistics*, 33:57–82.

William C Mann and Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Daniel Marcu. 2000. *The Theory and Practice of Discourse Parsing and Summarization*. MIT press.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830.

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 Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Rashmi Prasad, Bonnie Webber, and Alan Lee. 2018. Discourse annotation in the PDTB: The next generation. In *Proceedings 14th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, pages 87–97, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Valentina Pyatkin, Ayal Klein, Reut Tsarfaty, and Ido Dagan. 2020. QADiscourse - Discourse Relations as QA Pairs: Representation, Crowdsourcing and Baselines. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2804–2819, Online. Association for Computational Linguistics.

Ines Rehbein, Merel Scholman, and Vera Demberg. 2016. Annotating discourse relations in spoken language: A comparison of the PDTB and CCR frameworks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1039–1046, Portorož, Slovenia. European Language Resources Association (ELRA).

Charlotte Roze, Chloé Braud, and Philippe Muller. 2019. Which aspects of discourse relations are hard to learn? primitive decomposition for discourse relation classification. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 432–441, Stockholm, Sweden. Association for Computational Linguistics.

Ted J. M. Sanders, Wilbert P. M. Spooren, and Leo G. M. Noordman. 1993. Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics*, 4(2):93–134.

Ted JM Sanders, Vera Demberg, Jet Hoek, Merel CJ Scholman, Fatemeh Torabi Asr, Sandrine Zufferey, and Jacqueline Evers-Vermeul. 2018. Unifying dimensions in coherence relations: How various annotation frameworks are related. *Corpus Linguistics and Linguistic Theory*.

Ted JM Sanders, Wilbert PM Spooren, and Leo GM Noordman. 1992. Toward a taxonomy of coherence relations. *Discourse processes*, 15(1):1–35.

Tatjana Scheffler and Manfred Stede. 2016. Mapping PDTB-style connective annotation to RST-style discourse annotation. In *Proceedings of the 13th Conference on Natural Language Processing*, pages 242–247.

Wei Shi and Vera Demberg. 2019. Next sentence prediction helps implicit discourse relation classification within and across domains. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5790–5796, Hong Kong, China. Association for Computational Linguistics.

Karin Sim Smith. 2017. On integrating discourse in machine translation. In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pages 110–121, Copenhagen, Denmark. Association for Computational Linguistics.

Zirui Wang, Zihang Dai, Barnabás Póczos, and Jaime Carbonell. 2019. Characterizing and avoiding negative transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11293–11302.

Florian Wolf and Edward Gibson. 2004. Representing discourse coherence: A corpus-based analysis. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, pages 134–140, Geneva, Switzerland. COLING.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Amir Zeldes, Yang Janet Liu, Mikel Iruskieta, Philippe Muller, Chloé Braud, and Sonia Badene. 2021. The DISRPT 2021 shared task on elementary discourse unit segmentation, connective detection, and relation classification. In *Proceedings of the 2nd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2021)*, pages 1–12, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Zheng Zhao and Bonnie Webber. 2021. Revisiting shallow discourse parsing in the PDTB-3: Handling intrasentential implicits. In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*, pages 107–121, Punta Cana, Dominican Republic and Online. Association for Computational Linguistics.

A RST to UniDim Dimension Mapping Table

Table 9 shows the mapping of RST-DT relation labels to UniDim dimensions.

Class	End label	Nuc.	N-S	Pol.	Basic Op.	Impl. order	SoC	Temp.	Add. features
Background	Background	Mono	N-S	pos/neg	add	N.A.	obj	anti/N.A.	
	Background	Mono	S-N	pos/neg	add	N.A.	obj	chron/N.A.	
	Circumstance	Mono		pos/neg	add	N.A.	obj	syn/N.A.	
Cause	Cause	Mono	N-S	pos	cau	bas	obj	chron	
	Cause	Mono	S-N	pos	cau	non-b	obj	anti	
	Cause-result	Multi		pos	cau	bas/non-b	obj	chron/anti	
	Result	Mono	N-S	pos	cau	non-b	obj	anti	
	Result	Mono	S-N	pos	cau	bas	obj	chron	
	Consequence-n	Mono	N-S	pos	cau	non-b	obj	anti	
	Consequence-n	Mono	S-N	pos	cau	bas	obj	chron	
	Consequence-s	Mono	N-S	pos	cau	bas	obj	chron	
	Consequence-s	Mono	S-N	pos	cau	non-b	obj	anti	
	Consequence	Multi		pos	cau	bas/non-b	obj	chron/anti	
Comparison	Comparison	Both		pos	add	N.A.	obj/sub	N.A.	
	Preference	Mono		neg	add	N.A.	obj/sub	N.A.	
	Analogy	Both		pos	add	N.A.	sub	N.A.	
	Proportion	Multi		pos	add/cau	any	obj/sub	any	
Conditional	Condition	Mono	N-S	pos/neg	cau	non-b	obj/sub	anti/N.A.	conditional
	Condition	Mono	S-N	pos/neg	cau	bas	obj/sub	chron/N.A.	conditional
	Hypothetical	Mono	N-S	pos	cau	non-b	sub	N.A.	conditional
	Hypothetical	Mono	S-N	pos	cau	bas	sub	N.A.	conditional
	Contingency	Mono	N-S	pos/neg	cau	non-b	obj	anti	conditional
	Contingency	Mono	S-N	pos/neg	cau	bas	obj	chron	conditional
	Otherwise	Mono	N-S	neg	cau	bas	obj/sub	chron/N.A.	conditional
	Otherwise	Multi		neg	cau	bas	obj/sub	chron/N.A.	conditional
Contrast	Contrast	Multi		neg	add	N.A.	obj/sub	any	
	Concession	Mono	N-S	neg	cau	non-b	obj/sub	anti/N.A.	
	Concession	Mono	S-N	neg	cau	bas	obj/sub	chron/N.A.	
	Antithesis	Mono		neg	add/cau	any	obj/sub	any	
Elaboration	El.-additional	Mono		pos	add	N.A.	obj/sub	N.A.	
	El.-gen.-spec.	Mono		pos	add	N.A.	obj/sub	N.A.	specificity
	El.-part-whole	Mono		pos	add	N.A.	obj	N.A.	specificity
	El.-process-step	Mono		pos	add	N.A.	obj	N.A.	specificity
	El.-object-attr.	Mono		pos	add	N.A.	obj	N.A.	specificity
	El.-set-member	Mono		pos	add	N.A.	obj	N.A.	spec.-ex.
	Example	Mono		pos	add	N.A.	obj	N.A.	spec.-ex.
	Definition	Mono		pos	add	N.A.	obj	N.A.	specificity
Enablement	Purpose	Mono	N-S	pos	cau	bas	obj/sub	chron/N.A.	goal
	Purpose	Mono	S-N	pos	cau	non-b	obj/sub	anti/N.A.	goal
	Enablement	Mono	N-S	pos	cau	non-b	obj/sub	anti/N.A.	goal
	Enablement	Mono	S-N	pos	cau	bas	obj/sub	chron/N.A.	goal
Evaluation	Evaluation	Both		pos	add/cau	any	sub	N.A.	specificity
	Interpretation	Both		pos	add/cau	any	sub	N.A.	specificity
	Conclusion	Mono	N-S	pos	cau	bas	sub	N.A.	specificity
	Conclusion	Mono	S-N	pos	cau	non-b	sub	N.A.	specificity
	Conclusion	Multi		pos	cau	bas/non-b	sub	N.A.	specificity
	Comment	Mono		pos	add	N.A.	sub	N.A.	specificity
Explanation	Evidence	Mono	N-S	pos	cau	non-b	sub	anti	
	Evidence	Mono	S-N	pos	cau	bas	sub	chron	
	Exp.-argument.	Mono	N-S	pos	cau	non-b	obj	anti	
	Exp.-argument.	Mono	S-N	pos	cau	bas	obj	chron	
	Reason	Mono	N-S	pos	cau	non-b	obj	anti	
	Reason	Mono	S-N	pos	cau	bas	obj	chron	
	Reason	Multi		pos	cau	bas/non-b	obj	chron/anti	
Joint	List	Multi		pos	add	N.A.	obj/sub	syn/chron/N.A.	list
	Disjunction	Multi		pos/neg	add	N.A.	obj/sub	syn/N.A.	alternative
Summary	Summary	Mono		pos	add	N.A.	obj	N.A.	specificity
	Restatement	Mono		pos	add	N.A.	obj	N.A.	spec.-equiv.
Temporal	Temp.-before	Mono	N-S	pos	add	N.A.	obj	chron	
	Temp.-before	Mono	S-N	pos	add	N.A.	obj	anti	
	Temp.-after	Mono	N-S	pos	add	N.A.	obj	anti	
	Temp.-after	Mono	S-N	pos	add	N.A.	obj	chron	
	Temp.-same-time	Both		pos	add	N.A.	obj	syn	
	Sequence	Multi		pos	add	N.A.	obj	chron	
	Inverted-seq.	Multi		pos	add	N.A.	obj	anti	
Manner-Means	Means	Mono	N-S	pos	cau	non-b	obj	anti	
	Means	Mono	S-N	pos	cau	bas	obj	chron	goal
Topic-Comment	Problem-sol.-n	Mono	N-S	pos	cau	non-b	obj/sub	anti/N.A.	goal
	Problem-sol.-n	Mono	S-N	pos	cau	bas	obj/sub	chron/N.A.	goal
	Problem-sol.-s	Mono	N-S	pos	cau	bas	obj/sub	chron/N.A.	goal
	Problem-sol.-s	Mono	S-N	pos	cau	non-b	obj/sub	anti/N.A.	goal
	Problem-sol.	Multi		pos	cau	bas/non-b	obj/sub	achron/anti/N.A.	goal

Table 9: Mapping of RST relations to UniDim dimensions, taken from Sanders et al. (2018)

Table 9 is the mapping table of relation labels of RST-DT to UniDim dimensions. ***Nuc.*** means the nuclearity of a relation. ***N-S*** means whether the nuclearity is Nucleus-Satellite (N-S) or Satellite-Nucleus (S-N) or Nucleus-Nucleus (N-N). ***Pol.***, ***Basic Op.***, ***Impl. order***, ***Basic Op.***, ***SoC***, ***Temp.***, and ***Add. features*** denote polarity, basic operation, source of coherence, temporality and additional features, respectively.

B Relation Labels of PDTB 3.0 to UniDim Dimension Mapping Table

Table 10 shows the mapping of relation labels of PDTB 3.0 to UniDim dimensions.

Class_type	End label	A1-A2	Pol.	Basic Op.	Impl. order	SoC	Temp.	Add. features
Temporal								
Synchronous			pos	add	N.A.	obj	sync	
Asynchronous	Precedence	A1-A2	pos	add	N.A.	obj	chron	
	Precedence	A2-A1	pos	add	N.A.	obj	anti	
	Succession	A1-A2	pos	add	N.A.	obj	anti	
	Succession	A2-A1	pos	add	N.A.	obj	chron	
Contingency								
Cause	Reason	A1-A2	pos	cau	non-b	obj	anti	
	Reason	A2-A1	pos	cau	bas	obj	chron	
	Result	A1-A2	pos	cau	bas	obj	chron	goal
	Result	A1-A2	pos	cau	bas	obj	chron	goal
	NegResult		neg	cau	bas	obj	chron	
Cause+Belief	Reason+Belief	A1-A2	pos	cau	non-b	sub	NS	
	Reason+Belief	A2-A1	pos	cau	bas	sub	NS	
	Result+Belief	A1-A2	pos	cau	bas	sub	NS	
	Result+Belief	A2-A1	pos	cau	non-b	sub	NS	
Cause +SpeechAct	Reason+SpeechAct	A1-A2	pos	cau	non-b	sub	NS	
	Reason+SpeechAct	A2-A1	pos	cau	bas	sub	NS	
	Result+SpeechAct	A1-A2	pos	cau	bas	sub	NS	
	Result+SpeechAct	A2-A1	pos	cau	non-b	sub	NS	
Purpose	arg1-as-goal	A1-A2	pos	cau	non-b	obj/sub	NS	goal
	arg1-as-goal	A2-A1	pos	cau	bas	obj/sub	NS	goal
	arg2-as-goal	A1-A2	pos	cau	bas	sub	NS	goal
Condition	arg1-as-cond	A1-A2	pos	cau	bas	obj/sub	NS	conditional
	arg1-as-cond	A2-A1	pos	cau	non-b	obj/sub	NS	conditional
	arg2-as-cond	A1-A2	pos	cau	non-b	obj/sub	NS	conditional
	arg2-as-cond	A2-A1	pos	cau	bas	obj/sub	NS	conditional
Condition +SpeechAct			pos	cau	bas	sub	NS	conditional
Negative -Condition	arg1-as-negcond	A1-A2	neg	cau	bas	sub	NS	conditional
	arg1-as-negcond	A2-A1	neg	cau	non-b	sub	NS	conditional
	arg2-as-negcond	A1-A2	neg	cau	non-b	sub	NS	conditional
	arg2-as-negcond	A2-A1	neg	cau	bas	sub	NS	conditional
Negative- Condition+ SpeechAct			neg	cau	bas	sub	NS	conditional
Comparison								
Concession	arg1-as-denier	A1-A2	neg	cau	non-b	obj/sub	NS	
	arg1-as-denier	A2-A1	neg	cau	bas	obj/sub	NS	
	arg2-as-denier	A1-A2	neg	cau	bas	obj/sub	NS	
	arg2-as-denier	A2-A1	neg	cau	non-b	obj/sub	NS	
Concession +SpeechAct			neg	cau	bas	sub	NS	
Contrast			neg	add	NA	obj	NS	
Similarity			pos	add	NA	obj	NS	
Expansion								
Conjunction			pos	add	NA	obj/sub	NS	
Disjunction			neg	add	NA	obj/sub	NS	alternative
Equivalence			pos	add	NA	obj/sub	NS	
Exception	arg1-as-excpt		neg	add	NA	obj/sub	NS	
	arg2-as-excpt		neg	add	NA	obj/sub	NS	
Instantiation	arg1-as-instance		pos	add	NA	obj/sub	NS	specificity
	arg2-as-instance		pos	add	NA	obj/sub	NS	specificity
Level-of-detail	arg1-as-detail		pos	add	NA	obj/sub	NS	specificity
	arg2-as-detail		pos	add	NA	obj/sub	NS	specificity
Manner	arg1-as-manner	A1-A2	pos	add	NA	obj/sub	NS	specificity
	arg2-as-manner		pos	add	NA	obj/sub	NS	specificity
Substitution	arg1-as-subst	A1-A2	neg	cau	bas	obj/sub	NS	
	arg1-as-subst	A2-A1	neg	cau	non-b	obj/sub	NS	
	arg2-as-subst	A1-A2	neg	cau	non-b	obj/sub	NS	
	arg2-as-subst	A2-A1	neg	cau	bas	obj/sub	NS	

Table 10: Mapping of relations labels of PDTB 3.0 to UniDim dimensions.

Table 10 is the mapping table of relation labels of PDTB 3.0 to UniDim dimensions. A1-A2 means Argument 1 precedes Argument 2 and A2-A1 means Argument 2 precedes Argument 1 in the original text. The abbreviations are interpreted in the same way as in Table 9.

C Distribution of UniDim dimensions in RST-DT and PDTB 3.0

Figure 1 shows distribution of the polarity, basic operation, implication order, source of coherence, temporality and additional dimensions used in this paper.

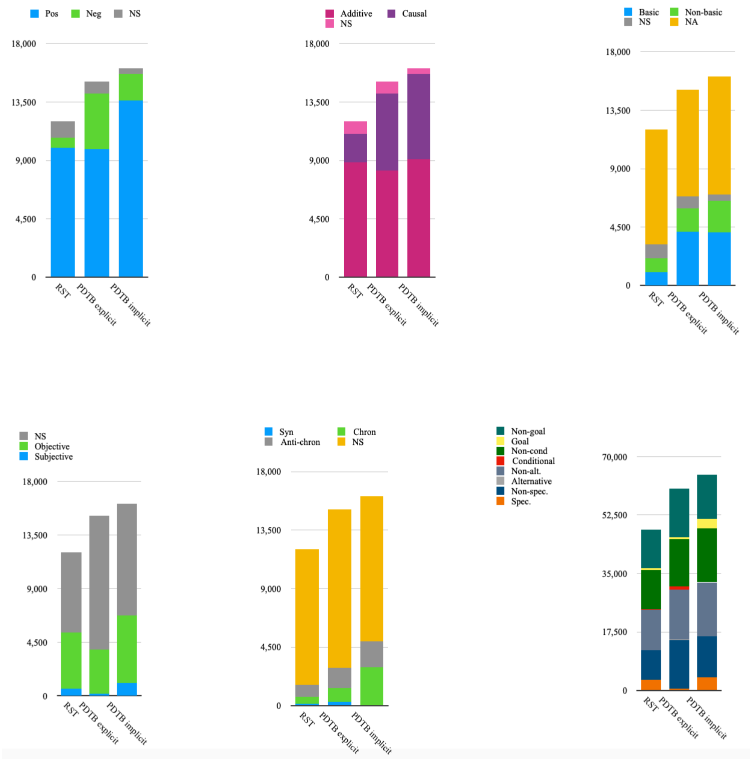

Figure 1: Distribution of the polarity, basic operation, and implication order dimensions (upper row, from left to right, respectively), and source of coherence, temporality and additional dimensions (lower row, from left to right, respectively) in the training sets of RST-DT and PDTB 3.0. We divide PDTB 3.0 based on explicit and implicit relation types.

D Hyper-parameters

For discourse relation classification described in section 3.1, the model is configured with a dropout rate of 0.2. The size of the output of the first MLP is set to 256 and the size of the second MLP output is 128. The model is trained with the AdamW optimizer (Loshchilov and Hutter, 2019), with a learning rate of $5e-5$. The batch size is set to 4 and the maximum norm of gradient clipping is set to 1. We use get_linear_schedule_with_warmup from the Transformers library as the learning rate scheduler. The maximum training epoch number is set to 10. The same setting is used in training the model for UniDim dimension prediction, the only exception being the learning rate, which is set to $1e-5$ to obtain good performance for this task.

For the cross-framework discourse relation classification task, the learning rate for transfer learning is $1e-5$ and as only parameters of the classifier layer are learnable, the maximum training epoch number is set to 50. The other hyper-parameters are the same as above.

We choose the best-performing model based on the performance at the validation set. The PyTorch library (Paszke et al., 2019) is used for implementation. The models are trained on an RTX2060 Super GPU.

The model for PDTB relation classification has 109,753,388 parameters and the training process took 6:25:23 (h:mm:ss) GPU hours for PDTB total relation classification, 2:56:58 GPU hours for PDTB explicit relation classification and 3:13:13 GPU hours for PDTB implicit relation classification. The model for RST relation classification has 109,494,544 parameters and the training process took 2:28:44 GPU hours. The number of parameters in the model for transfer learning is 2,064 and the training process took 4:38:43 GPU hours.

E Distribution of Relations in Training Data

Figures 2, 3, 4 and 5 shows the distribution of relations in the training sets used in the experiments, sorted in descending order.

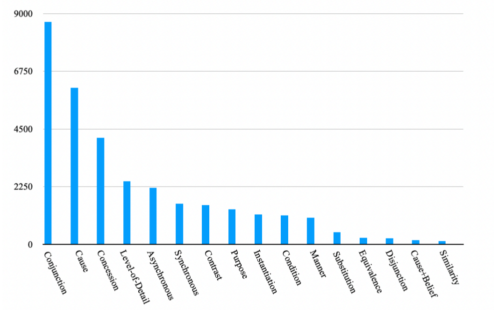

Figure 2: Distribution of PDTB relations in the experiment on PDTB where data of explicit and implicit relations are combined.

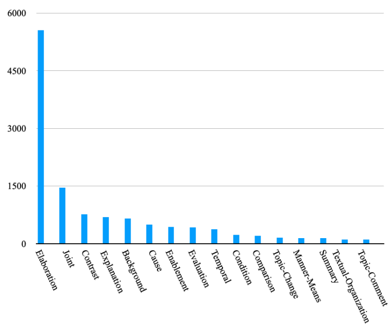

Figure 3: Distribution of RST relations in the training set.

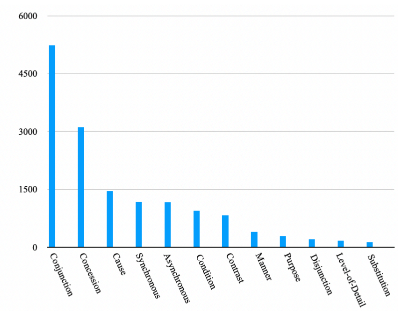

Figure 4: Distribution of PDTB explicit relations in the training set.

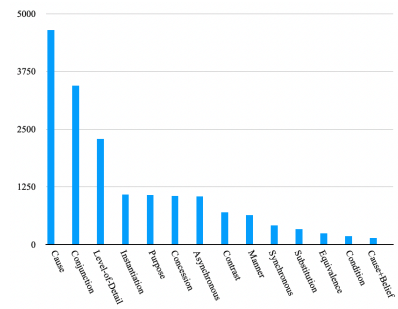

Figure 5: Distribution of PDTB implicit relations in the training set.

F PDTB Total Data Relation Classification

Table 11 shows the classification report on PDTB 3.0 (combining explicit and implicit relations) with BERT embeddings and UniDim dimensions as input features.

	Precision	Recall	F1	Support
Asynchronous	1.00	1.00	1.00	232
Cause	1.00	1.00	1.00	538
Cause+Belief	1.00	1.00	1.00	13
Concession	0.99	0.96	0.98	371
Condition	1.00	1.00	1.00	79
Conjunction	0.97	1.00	0.98	745
Contrast	1.00	1.00	1.00	102
Disjunction	1.00	1.00	1.00	20
Equivalence	0.00	0.00	0.00	25
Instantiation	0.00	0.00	0.00	117
Level-of-detail	0.00	0.00	0.00	202
Manner	0.07	0.96	0.14	26
Purpose	1.00	0.96	0.98	118
Similarity	0.00	0.00	0.00	12
Substitution	0.68	0.91	0.78	35
Synchronous	0.90	1.00	0.95	170
Accuracy		**0.86**		
Macro-F1	**0.66**	**0.74**	**0.67**	2805

Table 11: PDTB relation classification with BERT embeddings and UniDim dimensions as features.

Table 12 shows the classification report on PDTB 3.0 (combining explicit and implicit relations) with BERT embeddings as input.

	Precision	Recall	F1	Support
Asynchronous	0.79	0.65	0.71	232
Cause	0.71	0.62	0.66	538
Cause+Belief	0.00	0.00	0.00	13
Concession	0.78	0.83	0.80	371
Condition	0.92	0.87	0.90	79
Conjunction	0.71	0.85	0.77	745
Contrast	0.48	0.40	0.44	102
Disjunction	0.86	0.90	0.88	20
Equivalence	0.36	0.16	0.22	25
Instantiation	0.70	0.57	0.63	117
Level-of-detail	0.48	0.53	0.50	202
Manner	0.41	0.62	0.49	26
Purpose	0.87	0.84	0.85	118
Similarity	0.78	0.58	0.67	12
Substitution	0.53	0.49	0.51	35
Synchronous	0.74	0.64	0.68	170
Accuracy		**0.71**		
Macro-F1	**0.63**	**0.60**	**0.61**	2805

Table 12: PDTB relation classification with BERT embeddings as features.

G PDTB Explicit Relation Classification

Table 13 shows the classification report on PDTB 3.0 (explicit relations only) with BERT embeddings and UniDim dimensions as input features.

	Precision	Recall	F1	Support
Asynchronous	1.00	1.00	1.00	127
Cause	1.00	1.00	1.00	115
Concession	0.96	1.00	0.98	285
Condition	1.00	1.00	1.00	61
Conjunction	1.00	1.00	1.00	516
Contrast	1.00	1.00	1.00	50
Disjunction	1.00	1.00	1.00	18
Level-of-detail	0.00	0.00	0.00	20
Manner	0.35	1.00	0.52	11
Purpose	1.00	1.00	1.00	29
Substitution	0.00	0.00	0.00	13
Synchronous	1.00	1.00	1.00	126
Accuracy		**0.98**		
Macro-F1	**0.78**	**0.83**	**0.79**	1371

Table 13: Classification report of PDTB explicit relations with BERT embeddings and UniDim dimensions as features.

Table 14 shows the classification report on PDTB 3.0 (explicit relations only) with BERT embeddings as input features.

	Precision	Recall	F1	Support
Asynchronous	0.97	0.87	0.92	127
Cause	0.82	0.89	0.85	115
Concession	0.89	0.95	0.92	285
Condition	0.93	0.92	0.93	61
Conjunction	0.97	0.96	0.96	516
Contrast	0.52	0.48	0.50	50
Disjunction	0.90	1.00	0.95	18
Level-of-detail	0.71	0.75	0.73	20
Manner	0.42	0.91	0.57	11
Purpose	0.62	0.45	0.52	29
Substitution	1.00	0.92	0.96	13
Synchronous	0.81	0.71	0.76	126
Accuracy		**0.89**		
Macro-F1	**0.80**	**0.82**	**0.80**	1371

Table 14: Classification report of PDTB explicit relations with BERT embeddings as features.

H PDTB Explicit Relation Classification Ablation Studies

Table 15 shows the classification report on PDTB 3.0 (explicit relations only) with BERT embeddings and UniDim dimensions as input features, the polarity dimension being removed.

	Precision	Recall	F1	Support
Asynchronous	1.00	1.00	1.00	127
Cause	1.00	1.00	1.00	115
Concession	0.96	1.00	0.98	285
Condition	1.00	1.00	1.00	61
Conjunction	1.00	1.00	1.00	516
Contrast	0.62	1.00	0.76	50
Disjunction	1.00	1.00	1.00	18
Level-of-detail	0.00	0.00	0.00	20
Manner	0.35	1.00	0.52	11
Purpose	1.00	1.00	1.00	29
Substitution	0.00	0.00	0.00	13
Synchronous	1.00	0.75	0.86	126
Accuracy		**0.95**		
Macro-F1	**0.74**	**0.81**	**0.76**	1371

Table 15: Classification report of PDTB explicit relations, with the polarity dimension removed.

Table 16 shows the classification report on PDTB 3.0 (explicit relations only) with BERT embeddings and UniDim dimensions as input features, the basic operation dimension being removed.

	Precision	Recall	F1	Support
Asynchronous	1.00	1.00	1.00	127
Cause	1.00	1.00	1.00	115
Concession	0.96	1.00	0.98	285
Condition	1.00	1.00	1.00	61
Conjunction	1.00	1.00	1.00	516
Contrast	1.00	1.00	1.00	50
Disjunction	1.00	1.00	1.00	18
Level-of-detail	0.00	0.00	0.00	20
Manner	0.35	1.00	0.52	11
Purpose	1.00	1.00	1.00	29
Substitution	0.00	0.00	0.00	13
Synchronous	1.00	1.00	1.00	126
Accuracy		**0.98**		
Macro-F1	**0.78**	**0.83**	**0.79**	1371

Table 16: Classification report of PDTB explicit relations, with the basic operation dimension removed.

Table 17 shows the classification report on PDTB 3.0 (explicit relations only) with BERT embeddings and UniDim dimensions as input features, the source of coherence dimension being removed.

	Precision	Recall	F1	Support
Asynchronous	1.00	1.00	1.00	127
Cause	1.00	1.00	1.00	115
Concession	0.96	1.00	0.98	285
Condition	1.00	1.00	1.00	61
Conjunction	0.94	1.00	0.97	516
Contrast	0.74	1.00	0.85	50
Disjunction	0.00	0.00	0.00	18
Level-of-detail	0.00	0.00	0.00	20
Manner	0.35	1.00	0.52	11
Purpose	1.00	1.00	1.00	29
Substitution	0.00	0.00	0.00	13
Synchronous	1.00	0.75	0.86	126
Accuracy	**0.94**			
Macro-F1	**0.67**	**0.73**	**0.68**	1371

Table 17: Classification report of PDTB explicit relations, with the source of coherence dimension removed.

Table 18 shows the classification report on PDTB 3.0 (explicit relations only) with BERT embeddings and UniDim dimensions as input features, the implication order dimension being removed.

	Precision	Recall	F1	Support
Asynchronous	1.00	1.00	1.00	127
Cause	1.00	1.00	1.00	115
Concession	0.96	1.00	0.98	285
Condition	1.00	1.00	1.00	61
Conjunction	1.00	1.00	1.00	516
Contrast	1.00	1.00	1.00	50
Disjunction	1.00	1.00	1.00	18
Level-of-detail	0.00	0.00	0.00	20
Manner	0.35	1.00	0.52	11
Purpose	1.00	1.00	1.00	29
Substitution	0.00	0.00	0.00	13
Synchronous	1.00	1.00	1.00	126
Accuracy	**0.98**			
Macro-F1	**0.78**	**0.83**	**0.79**	1371

Table 18: Classification report of PDTB explicit relations, with the implication order dimension removed.

Table 19 shows the classification report on PDTB 3.0 (explicit relations only) with BERT embeddings and UniDim dimensions as input features, the temporality dimension being removed.

	Precision	Recall	F1	Support
Asynchronous	0.80	1.00	0.89	127
Cause	1.00	1.00	1.00	115
Concession	0.96	1.00	0.98	285
Condition	1.00	1.00	1.00	61
Conjunction	1.00	1.00	1.00	516
Contrast	1.00	1.00	1.00	50
Disjunction	1.00	1.00	1.00	18
Level-of-detail	0.00	0.00	0.00	20
Manner	0.35	1.00	0.52	11
Purpose	1.00	1.00	1.00	29
Substitution	0.00	0.00	0.00	13
Synchronous	1.00	0.75	0.86	126
Accuracy	**0.95**			
Macro-F1	**0.76**	**0.81**	**0.77**	1371

Table 19: Classification report of PDTB explicit relations, with the temporality dimension removed.

Table 20 shows the classification report on PDTB 3.0 (explicit relations only) with BERT embeddings and UniDim dimensions as input features, the additional dimensions being removed.

	Precision	Recall	F1	Support
Asynchronous	1.00	1.00	1.00	127
Cause	1.00	1.00	1.00	115
Concession	0.96	1.00	0.98	285
Condition	0.88	1.00	0.94	61
Conjunction	0.94	1.00	0.97	516
Contrast	1.00	1.00	1.00	50
Disjunction	1.00	1.00	1.00	18
Level-of-detail	0.00	0.00	0.00	20
Manner	0.00	0.00	0.00	11
Purpose	1.00	0.72	0.84	29
Substitution	0.00	0.00	0.00	13
Synchronous	1.00	1.00	1.00	126
Accuracy	**0.96**			
Macro-F1	**0.73**	**0.73**	**0.73**	1371

Table 20: Classification report of PDTB explicit relations, with the additional dimensions removed.

I PDTB Implicit Relation Classification

Table 21 shows the classification report on PDTB 3.0 (implicit relations only) with BERT embeddings and UniDim dimensions as input features.

	Precision	Recall	F1	Support
Asynchronous	1.00	1.00	1.00	95
Cause	1.00	1.00	1.00	366
Cause+Belief	1.00	0.42	0.59	12
Concession	1.00	0.92	0.96	84
Condition	1.00	1.00	1.00	12
Conjunction	0.90	1.00	0.95	221
Contrast	0.98	1.00	0.99	50
Equivalence	0.00	0.00	0.00	24
Instantiation	0.00	0.00	0.00	107
Level-of-detail	0.60	1.00	0.75	180
Manner	0.00	0.00	0.00	15
Purpose	0.92	0.94	0.93	88
Substitution	0.75	1.00	0.86	21
Synchronous	0.87	0.97	0.92	40
Accuracy	**0.87**			
Macro-F1	**0.72**	**0.73**	**0.71**	1315

Table 21: Classification report of implicit PDTB relations with BERT embeddings and UniDim dimensions as features.

Table 22 shows the classification report on PDTB 3.0 (implicit relations only) with only BERT embeddings as input features.

	Precision	Recall	F1	Support
Asynchronous	0.62	0.61	0.62	95
Cause	0.60	0.63	0.61	366
Cause+Belief	0.00	0.00	0.00	12
Concession	0.44	0.40	0.42	84
Condition	0.71	0.42	0.53	12
Conjunction	0.49	0.61	0.54	221
Contrast	0.45	0.42	0.43	50
Equivalence	0.12	0.04	0.06	24
Instantiation	0.77	0.54	0.64	107
Level-of-detail	0.45	0.48	0.46	180
Manner	0.38	0.60	0.46	15
Purpose	0.92	0.98	0.95	88
Substitution	0.43	0.48	0.45	21
Synchronous	0.27	0.10	0.15	40
Accuracy	**0.56**			
Macro-F1	**0.48**	**0.45**	**0.45**	1315

Table 22: Classification report of PDTB implicit relations with only BERT embeddings as features.

J PDTB Implicit Relation Classification Ablation Studies

Table 23 shows the classification report on PDTB 3.0 (implicit relations only) with BERT embeddings and UniDim dimensions as input features, the polarity dimension being removed.

	Precision	Recall	F1	Support
Asynchronous	1.00	1.00	1.00	95
Cause	1.00	1.00	1.00	366
Cause+Belief	1.00	0.42	0.59	12
Concession	0.96	0.92	0.94	84
Condition	1.00	0.75	0.86	12
Conjunction	0.90	1.00	0.95	221
Contrast	0.98	1.00	0.99	50
Equivalence	0.00	0.00	0.00	24
Instantiation	0.00	0.00	0.00	107
Level-of-detail	0.60	1.00	0.75	180
Manner	0.00	0.00	0.00	15
Purpose	0.92	0.94	0.93	88
Substitution	0.75	1.00	0.86	21
Synchronous	0.87	0.97	0.92	40
Accuracy		**0.87**		
Macro-F1	**0.71**	**0.71**	**0.70**	1315

Table 23: Classification report of PDTB implicit relations, with the polarity dimension removed.

Table 24 shows the classification report on PDTB 3.0 (implicit relations only) with BERT embeddings and UniDim dimensions as input features, the basic operation dimension being removed.

	Precision	Recall	F1	Support
Asynchronous	1.00	1.00	1.00	95
Cause	1.00	1.00	1.00	366
Cause+Belief	1.00	0.42	0.59	12
Concession	1.00	0.92	0.96	84
Condition	1.00	1.00	1.00	12
Conjunction	0.90	1.00	0.95	221
Contrast	1.00	1.00	1.00	50
Equivalence	0.00	0.00	0.00	24
Instantiation	0.00	0.00	0.00	107
Level-of-detail	0.60	1.00	0.75	180
Manner	0.00	0.00	0.00	15
Purpose	0.92	0.94	0.93	88
Substitution	0.75	1.00	0.86	21
Synchronous	0.87	0.97	0.92	40
Accuracy		**0.87**		
Macro-F1	**0.72**	**0.73**	**0.71**	1315

Table 24: Classification report of PDTB implicit relations, with the basic operation dimension removed.

Table 25 shows the classification report on PDTB 3.0 (implicit relations only) with BERT embeddings and UniDim dimensions as input features, the source of coherence dimension being removed.

	Precision	Recall	F1	Support
Asynchronous	1.00	1.00	1.00	95
Cause	1.00	1.00	1.00	366
Cause+Belief	1.00	0.42	0.59	12
Concession	1.00	0.92	0.96	84
Condition	1.00	1.00	1.00	12
Conjunction	0.90	1.00	0.95	221
Contrast	1.00	1.00	1.00	50
Equivalence	0.00	0.00	0.00	24
Instantiation	0.00	0.00	0.00	107
Level-of-detail	0.60	1.00	0.75	180
Manner	0.00	0.00	0.00	15
Purpose	0.92	0.94	0.93	88
Substitution	0.75	1.00	0.86	21
Synchronous	0.87	0.97	0.92	40
Accuracy		**0.87**		
Macro-F1	**0.72**	**0.73**	**0.71**	1315

Table 25: Classification report of PDTB implicit relations, with the source of coherence dimension removed. The result is the same as Table 24, where the basic operation dimension is removed.

Table 26 shows the classification report on PDTB 3.0 (implicit relations only) with BERT embeddings and UniDim dimensions as input features, the implication order dimension being removed.

	Precision	Recall	F1	Support
Asynchronous	1.00	1.00	1.00	95
Cause	1.00	1.00	1.00	366
Cause+Belief	0.00	0.00	0.00	12
Concession	0.80	1.00	0.89	84
Condition	1.00	1.00	1.00	12
Conjunction	0.90	1.00	0.95	221
Contrast	0.98	1.00	0.99	50
Equivalence	0.00	0.00	0.00	24
Instantiation	0.00	0.00	0.00	107
Level-of-detail	0.60	1.00	0.75	180
Manner	0.00	0.00	0.00	15
Purpose	0.87	0.94	0.91	88
Substitution	0.00	0.00	0.00	21
Synchronous	0.87	0.97	0.92	40
Accuracy		**0.86**		
Macro-F1	**0.57**	**0.64**	**0.60**	1315

Table 26: Classification report of PDTB implicit relations, with the implication order dimension removed.

Table 27 shows the classification report on PDTB 3.0 (implicit relations only) with BERT embeddings and UniDim dimensions as input features, the temporality dimension being removed.

	Precision	Recall	F1	Support
Asynchronous	0.99	1.00	0.99	95
Cause	1.00	1.00	1.00	366
Cause+Belief	1.00	0.42	0.59	12
Concession	1.00	0.92	0.96	84
Condition	1.00	1.00	1.00	12
Conjunction	0.90	1.00	0.95	221
Contrast	1.00	1.00	1.00	50
Equivalence	0.00	0.00	0.00	24
Instantiation	0.00	0.00	0.00	107
Level-of-detail	0.60	1.00	0.75	180
Manner	0.00	0.00	0.00	15
Purpose	0.92	0.94	0.93	88
Substitution	0.75	1.00	0.86	21
Synchronous	0.87	0.97	0.92	40
Accuracy		**0.87**		
Macro-F1	**0.72**	**0.73**	**0.71**	1315

Table 27: Classification report of PDTB implicit relations, with the temporality dimension removed.

Table 28 shows the classification report on PDTB 3.0 (implicit relations only) with BERT embeddings and UniDim dimensions as input features, the additional dimensions being removed.

	Precision	Recall	F1	Support
Asynchronous	0.99	1.00	0.99	95
Cause	1.00	1.00	1.00	366
Cause+Belief	1.00	0.42	0.59	12
Concession	0.96	0.92	0.94	84
Condition	1.00	0.75	0.86	12
Conjunction	0.40	1.00	0.58	221
Contrast	1.00	1.00	1.00	50
Equivalence	0.00	0.00	0.00	24
Instantiation	0.00	0.00	0.00	107
Level-of-detail	0.00	0.00	0.00	180
Manner	0.00	0.00	0.00	15
Purpose	0.92	0.94	0.93	88
Substitution	0.75	1.00	0.86	21
Synchronous	0.87	0.97	0.92	40
Accuracy		**0.73**		
Macro-F1	**0.64**	**0.64**	**0.62**	1315

Table 28: Classification report of PDTB implicit relations, with the additional dimensions removed.

K RST Relation Classification

Table 29 shows RST relation classification report with BERT embeddings and UniDim dimensions as input features.

Table 30 shows RST relation classification report with BERT embeddings as input features.

	Precision	Recall	F1	Support
Background	1.00	1.00	1.00	111
Cause	0.92	0.70	0.79	82
Comparison	0.00	0.00	0.00	29
Condition	1.00	1.00	1.00	48
Contrast	0.99	1.00	0.99	146
Elaboration	0.75	1.00	0.86	796
Enablement	0.92	1.00	0.96	46
Evaluation	0.99	1.00	0.99	80
Explanation	0.72	0.97	0.83	110
Joint	1.00	0.03	0.06	212
Manner-Means	0.00	0.00	0.00	27
Summary	0.00	0.00	0.00	32
Temporal	1.00	1.00	1.00	73
Textual-Organization	0.00	0.00	0.00	9
Topic-Change	0.28	1.00	0.44	13
Topic-Comment	0.71	0.21	0.32	24
Accuracy	**0.81**			
Macro-F1	**0.64**	**0.62**	**0.58**	1838

Table 29: RST relation classification report with BERT embeddings and UniDim dimensions as features.

	Precision	Recall	F1	Support
Background	0.47	0.35	0.40	111
Cause	0.50	0.17	0.25	82
Comparison	0.61	0.38	0.47	29
Condition	0.79	0.71	0.75	48
Contrast	0.75	0.68	0.72	146
Elaboration	0.65	0.88	0.75	796
Enablement	0.61	0.85	0.71	46
Evaluation	0.29	0.14	0.19	80
Explanation	0.46	0.27	0.34	110
Joint	0.67	0.62	0.64	212
Manner-Means	0.68	0.48	0.57	27
Summary	0.88	0.47	0.61	32
Temporal	0.74	0.27	0.40	73
Textual-Organization	0.44	0.44	0.44	9
Topic-Change	0.28	0.38	0.32	13
Topic-Comment	0.00	0.00	0.00	24
Accuracy	**0.63**			
Macro-F1	**0.55**	**0.44**	**0.47**	1838

Table 30: RST relation classification report using pre-trained BERT model.

Table 31 shows RST relation classification report using transfer learning from the PDTB relation classification model (combining PDTB explicit and implicit relation data during training) with BERT embeddings and UnDim dimensions as input features.

	Precision	Recall	F1	Support
Background	1.00	1.00	1.00	111
Cause	0.90	0.70	0.79	82
Comparison	0.00	0.00	0.00	29
Condition	1.00	0.98	0.99	48
Contrast	0.99	1.00	0.99	146
Elaboration	0.75	1.00	0.86	796
Enablement	0.92	1.00	0.96	46
Evaluation	1.00	1.00	1.00	80
Explanation	0.72	0.97	0.83	110
Joint	0.00	0.00	0.00	212
Manner-Means	0.00	0.00	0.00	27
Summary	0.00	0.00	0.00	32
Temporal	1.00	1.00	1.00	73
Textual-Organization	0.00	0.00	0.00	9
Topic-Change	0.28	1.00	0.44	13
Topic-Comment	0.71	0.21	0.32	24
Accuracy	**0.81**			
Macro-F1	**0.58**	**0.62**	**0.57**	1838

Table 31: Transfer learning for RST relation classification with the PDTB relation classification model with BERT embeddings and UniDim dimensions as input features.

Table 32 shows RST relation classification report using transfer learning from the pre-trained BERT model fine-tuned on PDTB relation classification task (combining PDTB explicit and implicit relation data).

	Precision	Recall	F1	Support
Background	0.51	0.27	0.35	111
Cause	0.17	0.07	0.10	82
Comparison	0.42	0.38	0.40	29
Condition	0.80	0.67	0.73	48
Contrast	0.75	0.73	0.74	146
Elaboration	0.60	0.82	0.69	796
Enablement	0.48	0.78	0.60	46
Evaluation	0.00	0.00	0.00	80
Explanation	0.40	0.15	0.22	110
Joint	0.57	0.66	0.61	212
Manner-Means	0.43	0.33	0.38	27
Summary	0.00	0.00	0.00	32
Temporal	0.53	0.36	0.43	73
Textual-Organization	0.00	0.00	0.00	9
Topic-Change	0.00	0.00	0.00	13
Topic-Comment	0.00	0.00	0.00	24
Accuracy	**0.58**			
Macro-F1	**0.35**	**0.33**	**0.33**	1838

Table 32: Transfer learning for RST relation classification using BERT embeddings as input.

L RST Relation Classification Ablation Studies

Table 33 shows the classification report on RST-DT with BERT embeddings and UniDim dimensions as input features, the polarity dimension being removed.

	Precision	Recall	F1	Support
Background	1.00	1.00	1.00	111
Cause	0.90	0.70	0.79	82
Comparison	0.00	0.00	0.00	29
Condition	1.00	0.94	0.97	48
Contrast	0.61	0.56	0.58	146
Elaboration	0.68	1.00	0.81	796
Enablement	0.92	1.00	0.96	46
Evaluation	1.00	0.57	0.73	80
Explanation	0.71	0.97	0.82	110
Joint	0.00	0.00	0.00	212
Manner-Means	0.00	0.00	0.00	27
Summary	0.00	0.00	0.00	32
Temporal	1.00	1.00	1.00	73
Textual-organization	0.00	0.00	0.00	9
Topic-Change	0.00	0.00	0.00	13
Topic-Comment	0.00	0.00	0.00	24
Accuracy	**0.74**			
Macro-F1	**0.49**	**0.48**	**0.48**	1838

Table 33: Classification report for RST, with the polarity dimension removed.

Table 34 shows the classification report on RST-DT with BERT embeddings and UniDim dimensions as input features, the basic operation dimension being removed.

	Precision	Recall	F1	Support
Background	0.95	1.00	0.97	111
Cause	0.90	0.70	0.79	82
Comparison	0.00	0.00	0.00	29
Condition	1.00	0.98	0.99	48
Contrast	0.99	1.00	0.99	146
Elaboration	0.73	1.00	0.84	796
Enablement	0.92	1.00	0.96	46
Evaluation	0.87	0.57	0.69	80
Explanation	0.72	0.97	0.83	110
Joint	0.00	0.00	0.00	212
Manner-Means	0.00	0.00	0.00	27
Summary	0.00	0.00	0.00	32
Temporal	1.00	1.00	1.00	73
Textual-Organization	0.00	0.00	0.00	9
Topic-Change	0.28	1.00	0.44	13
Topic-Comment	0.00	0.00	0.00	24
Accuracy	**0.78**			
Macro-F1	**0.52**	**0.58**	**0.53**	1838

Table 34: Classification report for RST, with the basic operation dimension removed.

Table 35 shows the classification report on RST-DT with BERT embeddings and UniDim dimensions as input features, the source of coherence dimension being removed.

	Precision	Recall	F1	Support
Background	0.95	1.00	0.97	111
Cause	0.84	0.70	0.76	82
Comparison	0.00	0.00	0.00	29
Condition	1.00	0.98	0.99	48
Contrast	0.99	1.00	0.99	146
Elaboration	0.73	1.00	0.84	796
Enablement	0.92	1.00	0.96	46
Evaluation	0.96	0.57	0.72	80
Explanation	0.72	0.97	0.83	110
Joint	0.00	0.00	0.00	212
Manner-Means	0.00	0.00	0.00	27
Summary	0.00	0.00	0.00	32
Temporal	1.00	1.00	1.00	73
Textual-Organization	0.00	0.00	0.00	9
Topic-Change	0.28	1.00	0.44	13
Topic-Comment	0.00	0.00	0.00	24
Accuracy		**0.78**		
Macro-F1	**0.52**	**0.58**	**0.53**	1838

Table 35: Classification report for RST, with the source of coherence dimension removed.

Table 36 shows the classification report on RST-DT with BERT embeddings and UniDim dimensions as input features, the implication order dimension being removed.

	Precision	Recall	F1	Support
Background	1.00	1.00	1.00	111
Cause	0.90	0.70	0.79	82
Comparison	0.00	0.00	0.00	29
Condition	1.00	0.98	0.99	48
Contrast	0.99	1.00	0.99	146
Elaboration	0.75	1.00	0.86	796
Enablement	0.84	1.00	0.91	46
Evaluation	0.99	1.00	0.99	80
Explanation	0.72	0.97	0.83	110
Joint	0.75	0.03	0.05	212
Manner-Means	0.00	0.00	0.00	27
Summary	0.00	0.00	0.00	32
Temporal	1.00	1.00	1.00	73
Textual-Organization	0.00	0.00	0.00	9
Topic-Change	0.28	1.00	0.44	13
Topic-Comment	0.00	0.00	0.00	24
Accuracy		**0.81**		
Macro-F1	**0.58**	**0.60**	**0.55**	1838

Table 36: Classification report for RST, with the implication order dimension removed.

Table 37 shows the classification report on RST-DT with BERT embeddings and UniDim dimensions as input features, the temporality dimension being removed.

	Precision	Recall	F1	Support
Background	1.00	1.00	1.00	111
Cause	0.92	0.70	0.79	82
Comparison	0.00	0.00	0.00	29
Condition	1.00	0.88	0.93	48
Contrast	0.99	1.00	0.99	146
Elaboration	0.75	1.00	0.86	796
Enablement	0.84	1.00	0.91	46
Evaluation	0.99	1.00	0.99	80
Explanation	0.69	0.97	0.81	110
Joint	1.00	0.03	0.06	212
Manner-Means	0.00	0.00	0.00	27
Summary	0.00	0.00	0.00	32
Temporal	1.00	1.00	1.00	73
Textual-Organization	0.00	0.00	0.00	9
Topic-Change	0.28	1.00	0.44	13
Topic-Comment	0.00	0.00	0.00	24
Accuracy		**0.80**		
Macro-F1	**0.59**	**0.60**	**0.55**	1838

Table 37: Classification report for RST, with the temporality dimension removed.

Table 38 shows the classification report on RST-DT with BERT embeddings and UniDim dimensions as input features, the additional dimensions being removed.

	Precision	Recall	F1	Support
Background	0.95	1.00	0.97	111
Cause	0.90	0.70	0.79	82
Comparison	0.00	0.00	0.00	29
Condition	1.00	0.81	0.90	48
Contrast	0.99	1.00	0.99	146
Elaboration	0.75	1.00	0.86	796
Enablement	0.84	1.00	0.91	46
Evaluation	0.90	1.00	0.95	80
Explanation	0.71	0.97	0.82	110
Joint	0.00	0.00	0.00	212
Manner-Means	0.00	0.00	0.00	27
Summary	0.00	0.00	0.00	32
Temporal	1.00	1.00	1.00	73
Textual-Organization	0.00	0.00	0.00	9
Topic-Change	0.28	1.00	0.44	13
Topic-Comment	0.00	0.00	0.00	24
Accuracy		**0.80**		
Macro-F1	**0.52**	**0.59**	**0.54**	1838

Table 38: Classification report for RST, with the additional dimensions removed.

M Cross-framework Discourse Relation Classification

Table 39 shows the classification report of the experiment using total PDTB data, where PDTB relation classification is the source task.

	P	R	$F1$	$P_{b.}$	$R_{b.}$	$F1_{b.}$	$C.$
Asynchronous	1.00	1.00	1.00	0.79	0.65	0.71	232
Cause	1.00	1.00	1.00	0.71	0.62	0.66	538
Cause+Belief	1.00	1.00	1.00	0.00	0.00	0.00	13
Concession	0.99	0.96	0.98	0.78	0.83	0.80	371
Condition	1.00	1.00	1.00	0.92	0.87	0.90	79
Conjunction	0.97	1.00	0.98	0.71	0.85	0.77	745
Contrast	1.00	1.00	1.00	0.48	0.40	0.44	102
Disjunction	1.00	1.00	1.00	0.86	0.90	0.88	20
Equivalence	0.00	0.00	0.00	0.36	0.16	0.22	25
Instantiation	0.00	0.00	0.00	0.70	0.57	0.63	117
Level-of-detail	0.00	0.00	0.00	0.48	0.53	0.50	202
Manner	0.07	0.96	0.14	0.41	0.62	0.49	26
Purpose	1.00	0.96	0.98	0.87	0.84	0.85	118
Similarity	0.00	0.00	0.00	0.78	0.58	0.67	12
Substitution	0.68	0.91	0.78	0.53	0.49	0.51	35
Synchronous	0.90	1.00	0.95	0.74	0.64	0.68	170
Acc.		**0.86**		0.71 (vs. DISRPT 2021: 0.74)			
Macro-F1	**0.66**	**0.74**	**0.67**	0.63	0.60	0.61	2805

Table 39: Results of relation classification on total PDTB data. Blue columns show our results and uncolored columns show results of the baseline model.

Table 40 shows the classification report of the target task, i.e. RST relation classification.

	P	R	$F1$	$P_{b.}$	$R_{b.}$	$F1_{b.}$	$C.$
Background	1.00	1.00	1.00	0.51	0.27	0.35	111
Cause	0.90	0.70	0.79	0.17	0.07	0.10	82
Comparison	0.00	0.00	0.00	0.42	0.38	0.40	29
Condition	1.00	0.98	0.99	0.80	0.67	0.73	48
Contrast	0.99	1.00	0.99	0.75	0.73	0.74	146
Elaboration	0.75	1.00	0.86	0.60	0.82	0.69	796
Enablement	0.92	1.00	0.96	0.48	0.78	0.60	46
Evaluation	1.00	1.00	1.00	0.00	0.00	0.00	80
Explanation	0.72	0.97	0.83	0.40	0.15	0.22	110
Joint	0.00	0.00	0.00	0.57	0.66	0.61	212
Manner-Means	0.00	0.00	0.00	0.43	0.33	0.38	27
Summary	0.00	0.00	0.00	0.00	0.00	0.00	32
Temporal	1.00	1.00	1.00	0.53	0.36	0.43	73
Textual-Organization	0.00	0.00	0.00	0.00	0.00	0.00	9
Topic-Change	0.28	1.00	0.44	0.00	0.00	0.00	13
Topic-Comment	0.71	0.21	0.32	0.00	0.00	0.00	24
Acc.		**0.81**			0.58		
Macro-F1	**0.58**	**0.62**	**0.57**	0.35	0.33	0.33	1838
RST acc		0.81			0.63		
RST Macro-F1	0.64	0.62	0.58	0.55	0.44	0.47	1838

Table 40: Results of the target task. The results of training a model specifically for RST relation classification with our method are shown in blue columns and the uncolored columns show results of the baseline model.