HITS at DISRPT 2023: Discourse Segmentation, Connective Detection, and Relation Classification

Wei Liu* and **Fan Yi*** and **Michael Strube**
Heidelberg Institute for Theoretical Studies
{wei.liu, yi.fan, michael.strube}@h-its.org

Abstract

HITS participated in the Discourse Segmentation (DS, Task 1) and Connective Detection (CD, Task 2) tasks at the DISRPT 2023. Task 1 focuses on segmenting the text into discourse units, while Task 2 aims to detect the discourse connectives. We deployed a framework based on different pre-trained models according to the target language for these two tasks.

HITS also participated in the Relation Classification track (Task 3). The main task was recognizing the discourse relation between text spans from different languages. We designed a joint model for languages with a small corpus while separate models for large corpora. The adversarial training strategy is applied to enhance the robustness of relation classifiers.

The implementation of our models for three tasks is available at https://github.com/liuwei1206/disrpt2023.

1 Task and Data

The 2023 shared task provides 3 sub-tasks, including discourse segmentation (DS, Task 1), Connective Detection (CD, Task 2), and Relation Classification (RC, Task 3).

Task 1 focuses on conducting discourse units segmentation under different formalisms, such as Rhetorical Structure Theory (RST, MANN and Thompson, 1988), Segmented Discourse Representation Theory (SDRT, Lascarides and Asher, 2007) and Penn Discourse Treebank (PDTB, Miltsakaki et al., 2004). As different corpora, languages and formalisms or theories use different segmentation guidelines, the challenge is to design flexible methods to deal with various situations. The aim of Task 2 is to identify discourse connectives in the text.

Relation classification aims to identify the discourse relation, such as *Cause* and *Comparison*, between two text spans. The shared task provides 26 corpora covering 13 languages, including Basque, Chinese, Dutch, English, French, German, Italian, Persia, Portugal, Russian, Spanish, Thai, and Turkish. Most of corpora are annotated with Rhetorical Structure Theory (RST, MANN and Thompson, 1988) and Penn Discourse Treebank (PDTB, Miltsakaki et al., 2004), with a small part using Segmented Discourse Representation Theory (SDRT, Lascarides and Asher, 2007) and Discourse Dependency Framework (DEP, Stede et al., 2016). We show the statistics of relation corpora in Table 2.

2 Discourse Segmentation and Connective Detection

2.1 Approach

Our framework for Task 1 and Task 2 is composed of a BERT-based model (Devlin et al., 2019), Bi-LSTM (Hochreiter and Schmidhuber, 1997) and conditional random field (CRF, Lafferty et al., 2001). In our framework, we first obtain the embedding of the input text via a BERT-based model. We then use Bi-LSTM to capture the contextual information and generate a richer contextual representation by exploiting the sequential relationships between words. Finally, CRF can globally optimise the label sequence based on the contextual information of the current word and the relationship between the preceding and following labels, resulting in better consistency and rationality of the predicted label sequence.

2.2 Experiments

As the BERT-based model provides the embedding for the input text, choosing an appropriate one according to the language is essential for the framework. We choose at least two pre-trained BERT-based models for each language and fine-tune the parameters to achieve the best performance for our framework. After several experiments and comparing different BERT-based models, our final choice for the BERT-based model for different languages

*Equal contribution. Wei is responsible for relation classification, while Yi works on discourse segmentation and connective detection.

Proceedings of the 3rd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2023), pages 43–49
July 14, 2023

Framework	Corpus	Task 1/2				Task3			
		Label	Train	Dev	Test	Label	Train	Dev	Test
RST	deu.rst.pcc (Stede and Neumann, 2014)	2	1773	207	213	26	2164	241	260
	eng.rst.gum (Zeldes, 2017)	2	9234	1221	1201	14	19497	2618	2576
	eng.rst.rstdt (Lynn Carlson, 2002; Carlson et al., 2003)	2	6672	717	929	17	16003	1622	2156
	eus.rst.ert (Iruskieta et al., 2013; Aranzabe et al., 2015)	2	1599	366	415	29	2534	679	615
	fas.rst.prstc (Shahmohammadi et al., 2021)	2	1713	202	264	17	4101	500	593
	nld.rst.nldt (Redeker et al., 2012)	2	1156	255	240	32	1609	332	326
	por.rst.cstn (Cardoso et al., 2011)	2	1825	257	139	32	8798	1286	1249
	rus.rst.rrt (Pisarevskaya et al., 2017; Toldova et al., 2017)	2	18932	2025	2087	22	28869	2856	2844
	spa.rst.rststb (da Cunha et al., 2011)	2	1548	254	287	28	2241	384	427
	spa.rst.sctb (Cao et al., 2018a, 2017b,a, 2016)	2	326	76	114	24	440	95	160
	zho.rst.sctb (Cao et al., 2018b, 2017c,a, 2016)	2	361	86	133	26	440	95	160
	zho.rst.gcdt (Peng et al., 2022)	2	2026	331	335	31	6455	1007	954
PDTB	eng.pdtb.pdtb (Webber et al., 2019)	3	44563	1703	2364	23	43920	1674	2257
	eng.pdtb.tedm (Zeyrek et al., 2018)	3	-	143	238	20	-	179	352
	ita.pdtb.luna (Tonelli et al., 2010)	2	3721	775	1315	15	957	211	382
	por.pdtb.crpc (Mendes and Lejeune, 2022)	3	4078	581	535	22	8798	1286	1249
	por.pdtb.tedm (Zeyrek et al., 2018)	3	-	148	246	20	-	191	365
	tha.pdtb.tdtb	3	5076	633	825	20	8279	1244	1345
	tur.pdtb.tdb (Zeyrek Bozşahin et al., 2013)	3	24960	2948	3289	23	2452	313	423
	tur.pdtb.tedm (Zeyrek et al., 2018)	3	-	141	269	23	-	214	365
	zho.pdtb.cdtb (Zhou et al., 2014)	3	2049	438	404	9	3657	855	758
SDRT	eng.sdrt.stac (Asher et al., 2016)	2	8754	991	1342	16	9581	1146	1511
	fra.sdrt.annodis (Afantenos et al., 2012)	2	1020	245	242	18	2186	529	626
DEP	eng.dep.covdtb (Nishida and Matsumoto, 2022)	2	-	1162	1181	12	-	2400	2587
	eng.dep.scidtb (Yang and Li, 2018)	2	2570	815	817	24	6061	1934	1912
	zho.dep.scidtb (Cheng and Li, 2019)	2	308	103	89	23	803	282	216

Table 1: Statistics of corpora provided by the shared task.

Language	Pre-trained model choice
deu	xlm-roberta-base
eng	roberta-base
eus	ixa-ehu/berteus-base-cased
fas	HooshvareLab/bert-fa-base-uncased
fra	xlm-roberta-base
ita	xlm-roberta-base
nld	pdelobelle/robbert-v2-dutch-base
por	neuralmind/bert-base-portuguese-cased
rus	DeepPavlov/rubert-base-cased
spa	dccuchile/bert-base-spanish-wwm-cased
tur	dbmdz/bert-base-turkish-cased
zho	bert-base-chinese
tha	airesearch/wangchanberta-base-att-spm-uncased

Table 2: Model choice for different languages

is shown in Table 2. Our framework is trained with batch size 16 for each corpus, and the maximum input sequence length is 512. If the input sequence length exceeds 512, then our framework will slice it into two or more segments. The maximum length of all segments is also 512. The LSTM in our framework has two layers, and both of them are bi-directional. The criterion that we choose those BERT-based models in our framework is their best performance with corresponding parameters. In addition, all the pre-trained models we use for this shared task are provided by HuggingFace*. The result of our framework's performance with golden treebanked data for Task 1 and Task 2 shows in Table 3. We use our trained model on another corpus to evaluate corpora that do not have a training corpus and select the best one, shown in Table 4.

However, due to the time limitation, we only tuned all pre-trained models and tested our framework with the golden treebanked data as the input. Besides, we observed that normally the larger model performs better than the base model. For instance, for the corpus eng.dep.scidtb, we use the best parameters we tuned for the Roberta-base model (Liu et al., 2019) for the Roberta-large model, our framework's performance will increase 0.28% and 0.11% in the development set and test set separately. Also, we tried the Adversarial Training strategy (Miyato et al., 2016) and the Bootstrap aggregating strategy (Breiman, 1996), which is a commonly used ensemble learning method, separately with our framework. We only test the Bootstrap aggregating strategy on the corpus ita.pdtb.luna. We use the best and second-best learning rates on training with our framework to generate two models first. Then, we change the xlm-base model to dbmdz/bert-base-italian-uncased, and also apply the best and second-best learning rates to generate two trained models. Every time when we train these models, we tune the

*https://huggingface.co/

Corpus	F_1
deu.rst.pcc	96.19%
eng.rst.gum	81.22%
eng.rst.rstdt	97.36%
eus.rst.ert	89.85%
fas.rst.prstc	93.05%
nld.rst.nldt	93.64%
por.rst.cstn	94.63%
rus.rst.rrt	85.05%
spa.rst.rststb	90.87%
spa.rst.sctb	83.17%
zho.rst.sctb	80.12%
zho.rst.gcdt	91.37%
eng.pdtb.pdtb	93.47%
ita.pdtb.luna	66.41%
por.pdtb.crpc	79.74%
tha.pdtb.tdtb	86.92%
tur.pdtb.tdb	84.89%
zho.pdtb.cdtb	87.40%
eng.sdrt.stac	95.84%
fra.sdrt.annodis	88.45%
eng.dep.scidtb	94.97%
zho.dep.scidtb	90.59%
Mean	88.41%

Table 3: Results of Task 1 and Task 2 for corpora with a training dataset

ratio of the corpus to 1 means we use the whole corpus. Note that we can tune the ratio to sample randomly the percentage of data from the training dataset. Then we let all models vote in the development set and test set. Finally, we tally the results of the voting to determine the final model predictions. In our experiment setting, we follow the majority vote, which implies every vote contributes equally to the final result and the most voted result is selected. We found this simple setting of Bootstrap aggregating strategy can improve the F1 score by 0.16% and 0.13% on the development set and test set respectively. During the test of the Adversarial Training strategy, we only test on a few corpora. The result shows in Table 5. We observe that the performance increases in almost all corpora we test, which means this strategy functions. Insufficient time prevented us from exploring the result for the Adversarial Training strategy, the Bootstrap aggregating strategy, and the larger pre-trained models separately and in combinations of them on all corpora with plain text input and golden treebanked input settings.

Corpus	F_1	model source
eng.pdtb.tedm	78.56%	eng.pdtb.pdtb
por.pdtb.tedm	80.19%	por.pdtb.crpc
tur.pdtb.tedm	66.15%	tur.pdtb.tdb
eng.dep.covdtb	90.14%	eng.dep.scidtb
Mean	78.76%	-

Table 4: Results of Task 1 and Task 2 for corpora without a training dataset. The models used for generating the result are trained on is noted in the model source column.

Corpus	F_1	vs. without adv
deu.rst.pcc	96.59%	+0.40%
eng.sdrt.stac	97.21%	−0.15%
eus.rst.ert	90.10%	+0.25%
fas.rst.prstc	93.14%	+0.09%
nld.rst.nldt	96.46%	+2.82%
por.rst.cstn	95.85%	+1.22%
spa.rst.rststb	91.02%	+0.15%
spa.rst.sctb	83.76%	+0.59%
Mean	93.02%	+0.67125%

Table 5: Comparison between applying Adversarial Training strategy and without it for our framework on the dataset we have tested.

3 Relation Classification

3.1 Approach

The relation classifiers employed in this work follow an architecture widely used for text classification tasks: pre-trained models as the encoder and a linear network as the classification layer. The training of classifiers on each corpus varies from each other depending on the corpus size. Specifically, we train individual classifiers for large corpora (e.g., eng.pdtb.pdtb) but a joint model for a set of small datasets. This is because a large number of instances is sufficient to train a good classifier, while a small corpus can lead to underfitting.

For large corpora, including eng.rst.gum, eng.rst.rstdt, eus.rst.ert, zho.rst.gcdt, eng.pdtb.pdtb, eng.sdrt.stac, and fra.sdrt.annodis, individual classifiers are trained for them. For small corpora, we divide them into three groups according their annotation framework. The first is the **RST-group**, containing deu.rst.pcc, fas.rst.prstc, nld.rst.nldt, por.rst.cstn, rus.rst.rrt, spa.rst.rststb, spa.rst.sctb, and zho.rst.sctb. We train a joint model called **joint-RST** on the RST-group corpus. The second is the **PDTB-group**, covering ita.pdtb.luna,

Model type	Corpus	Encoder
individual	eng.rst.gum eng.rst.rstdt eng.pdtb.pdtb eng.sdrt.stac	roberta-large
	eus.rst.ert	berteus-base-cased
	zho.rst.gcdt zho.pdtb.cdtb	macbert-large
	fra.sdrt.annodis	camembert-large
joint-RST	deu.rst.pcc fas.rst.prstc nld.rst.nldt por.rst.cstn rus.rst.rrt spa.rst.rststb spa.rst.sctb zho.rst.sctb	xlm-roberta-large
joint-PDTB	eng.pdtb.tedm ita.pdtb.luna por.pdtb.crpc por.pdtb.tedm tha.pdtb.tdtb tur.pdtb.tdb tur.pdtb.tedm	
joint-DEP	eng.dep.scidtb eng.dep.covdtb zho.dep.scidtb	

Table 6: Training strategy for different relation corpora. "individual" means training a corpus-specific model.

por.pdtb.crpc, tha.pdtb.tdtb, and tur.pdtb.tdb, and the joint model **joint-PDTB** is trained on this group. The last is the **DEP-group**, including eng.dep.scidtb and zho.dep.scidtb, and its corresponding model is **joint-DEP**.

During training, adversarial strategy (Miyato et al., 2016) is applied to improve the robustness of classifiers. For corpora without a training set, we evaluate them with the joint model of the corresponding annotation framework. We summarize the setup for each corpus in Table 6.

3.2 Experiments

We train relation classifiers based on the corpora provided by the shared task. During the evaluation, we report the result of a model on the test set using the checkpoint that achieves the best performance on the development set.

Table 7 shows the results on corpora with training sets. We find that small corpora can significantly benefit from joint training. For example, the joint-RST outperforms a relation classifier trained on deu.rst.pcc solely more than 5 points (i.e., 30.00% → 35.77%). However, the joint model performs worse on large corpora, such as eng.rst.gum, decreasing the accuracy from 65.67% to 62.06%, compared to the individual model. Due to time constraints, we can not finish the ablation study on all corpora.

Model type	Corpus	Accuracy
Individual	eng.rst.gum	65.67
	eng.rst.rstdt	66.40
	eng.pdtb.pdtb	74.75
	eng.sdrt.stac	62.85
	eus.rst.ert	56.64
	zho.rst.gcdt	56.14
	zho.pdtb.pdtb	85.36
	fra.sdrt.annodis	50.08
joint-RST	deu.rst.pcc	35.77
	fas.rst.prstc	55.91
	nld.rst.nldt	55.69
	por.rst.cstn	68.38
	rus.rst.rrt	62.05
	spa.rst.rststb	58.69
	spa.pdtb.crpc	64.15
	zho.rst.sctb	62.26
joint-PDTB	ita.pdtb.luna	67.89
	por.pdtb.crpc	77.80
	tha.pdtb.tdtb	96.80
	tur.pdtb.tdb	56.64
joint-DEP	eng.dep.scidtb	75.30
	zho.dep.scidtb	67.44
	Mean	64.67

Table 7: Results (Task 3) for corpora with a training set.

Corpus	Accuracy
eng.pdtb.tedm	65.53
por.pdtb.tedm	67.03
tur.pdtb.tedm	56.87
eng.dep.covdtb	70.03
Mean	64.87

Table 8: Results (Task 3) for corpora without a training set.

The shared task also provides evaluation corpora without training sets. The primary goal is to test the zero-shot performance of trained classifiers. Our joint models are well suited for this setting since they have a large label set, inheriting from a group of small corpora. Table 8 shows joint models' results on those evaluation corpora. Surprisingly, our joint models perform well under the zero-shot setting, achieving an average accuracy of 64.87%, close to the performance of corpora with a training set (i.e., 64.67%).

4 Conclusion

In this paper, we present our models in the shared task DISRPT 2023. For Tasks 1 and 2, we employ a pre-trained model+BiLSTM+CRF to capture textual information and dependency between successive labels. For Task 3, we design a joint training strategy for small corpora, which can compensate for underfitting caused by limited training instances.

Acknowledgements

The authors would like to thank the anonymous reviewers for their comments. This work has been funded by the Klaus Tschira Foundation, Heidelberg, Germany.

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).

María Jesús Aranzabe, Aitziber Atutxa, Kepa Bengoetxea, Arantza Díaz, Deliana Ilarraza, Iakes Goenaga, and Koldo Gojenola. 2015. Automatic conversion of the basque dependency treebank to universal dependencies. In *the fourteenth international workshop on treebanks an linguistic theories (TLT14)*, pages 233–241, Warsaw, Poland.

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).

Leo Breiman. 1996. Bagging predictors. *Machine Learning*, 24(2):123–140.

Shuyuan Cao, Iria da Cunha, and Mikel Iruskieta. 2016. A corpus-based approach for Spanish-Chinese language learning. In *Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA2016)*, pages 97–106, Osaka, Japan. The COLING 2016 Organizing Committee.

Shuyuan Cao, Iria Da-Cunha, and Mikel Iruskieta. 2017a. Toward the elaboration of a spanish-chinese parallel annotated corpus. In *Professional and Academic Discourse: an Interdisciplinary Perspective*, volume 2 of *EPiC Series in Language and Linguistics*, pages 315–324. EasyChair.

Shuyuan Cao, Iria da Cunha, and Mikel Iruskieta. 2018a. The RST Spanish-Chinese treebank. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 156–166, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Shuyuan Cao, Iria da Cunha, and Mikel Iruskieta. 2018b. The RST Spanish-Chinese treebank. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 156–166, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Shuyuan Cao, Nianwen Xue, Iria da Cunha, Mikel Iruskieta, and Chuan Wang. 2017b. Discourse segmentation for building a RST Chinese treebank. In *Proceedings of the 6th Workshop on Recent Advances in RST and Related Formalisms*, pages 73–81, Santiago de Compostela, Spain. Association for Computational Linguistics.

Shuyuan Cao, Nianwen Xue, Iria da Cunha, Mikel Iruskieta, and Chuan Wang. 2017c. Discourse segmentation for building a RST Chinese treebank. In *Proceedings of the 6th Workshop on Recent Advances in RST and Related Formalisms*, pages 73–81, Santiago de Compostela, Spain. Association for Computational Linguistics.

Paula Christina Figueira Cardoso, Erick Galani Maziero, Maria Lucía del Rosario Castro Jorge, M. Eloize, R. Kibar Aji Seno, Ariani Di Felippo, Lucia Helena Machado Rino, Maria das Graças Volpe Nunes, and Thiago Alexandre Salgueiro Pardo. 2011. CSTNews - a discourse-annotated corpus for single and multi-document summarization of news texts in Brazilian Portuguese. In *Proceedings of the 3rd RST Brazilian Meeting*, pages 88–105, Cuiabá, Brazil.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2003. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In *Current and New Directions in Discourse and Dialogue*, Text, Speech and Language Technology 22, pages 85–112. Kluwer, Dordrecht.

Yi Cheng and Sujian Li. 2019. Zero-shot Chinese discourse dependency parsing via cross-lingual mapping. In *Proceedings of the 1st Workshop on Discourse Structure in Neural NLG*, pages 24–29, Tokyo, Japan. Association for Computational Linguistics.

Iria da Cunha, Juan-Manuel Torres-Moreno, and Gerardo Sierra. 2011. On the development of the RST Spanish treebank. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 1–10, Portland, Oregon, USA. Association for Computational Linguistics.

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.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.*, 9(8):1735–1780.

Mikel Iruskieta, María Jesús Aranzabe, Arantza Diaz de Ilarraza, Itziar Gonzalez-Dios, Mikel Lersundi, and Oier Lopez de Lacalle. 2013. The RST Basque TreeBank: An online search interface to check rhetorical relations. In *4th Workshop on RST and Discourse Studies*, pages 40–49, Fortaleza, Brasil.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, page 282–289, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Alex Lascarides and Nicholas Asher. 2007. *Segmented Discourse Representation Theory: Dynamic Semantics With Discourse Structure*, volume 3, pages 87–124.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Mary Ellen Okurowski Lynn Carlson, Daniel Marcu. 2002. RST Discourse Treebank LDC2002T07.

WILLIAM MANN and Sandra Thompson. 1988. Rethorical structure theory: Toward a functional theory of text organization. *Text*, 8:243–281.

Amália Mendes and Pierre Lejeune. 2022. Crpc-db a discourse bank for portuguese. In *Computational Processing of the Portuguese Language: 15th International Conference, PROPOR 2022, Fortaleza, Brazil, March 21–23, 2022, Proceedings*, pages 79–89. Springer.

Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. 2004. The Penn Discourse Treebank. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Takeru Miyato, Andrew M Dai, and Ian Goodfellow. 2016. Adversarial training methods for semi-supervised text classification. *arXiv preprint arXiv:1605.07725*.

Noriki Nishida and Yuji Matsumoto. 2022. Out-of-domain discourse dependency parsing via bootstrapping: An empirical analysis on its effectiveness and limitation. *Transactions of the Association for Computational Linguistics*, 10:127–144.

Siyao Peng, Yang Janet Liu, and Amir Zeldes. 2022. Chinese Discourse Annotation Reference Manual. Research Report, Georgetown University (Washington, D.C.).

Dina Pisarevskaya, Margarita Ananyeva, Maria Kobozeva, A. Nasedkin, S. Nikiforova, I. Pavlova, and A. Shelepov. 2017. Towards building a discourse-annotated corpus of russian. In *Computational Linguistics and Intellectual Technologies: 23rd International Conference on Computational Linguistics and Intellectual Technologies "Dialogue"*, pages 194–204.

Gisela Redeker, Ildikó Berzlánovich, Nynke van der Vliet, Gosse Bouma, and Markus Egg. 2012. Multi-layer discourse annotation of a Dutch text corpus. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2820–2825, Istanbul, Turkey. European Language Resources Association (ELRA).

Sara Shahmohammadi, Hadi Veisi, and Ali Darzi. 2021. Persian Rhetorical Structure Theory. *arXiv preprint arXiv:2106.13833*.

Manfred Stede, Stergos Afantenos, Andreas Peldszus, Nicholas Asher, and Jérémy Perret. 2016. Parallel discourse annotations on a corpus of short texts. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1051–1058, Portorož, Slovenia. European Language Resources Association (ELRA).

Manfred Stede and Arne Neumann. 2014. Potsdam commentary corpus 2.0: Annotation for discourse research. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 925–929, Reykjavik, Iceland. European Language Resources Association (ELRA).

Svetlana Toldova, Dina Pisarevskaya, Margarita Ananyeva, Maria Kobozeva, Alexander Nasedkin, Sofia Nikiforova, Irina Pavlova, and Alexey Shelepov. 2017. Rhetorical relations markers in Russian RST treebank. In *Proceedings of the 6th Workshop on Recent Advances in RST and Related Formalisms*, pages 29–33, Santiago de Compostela, Spain. Association for Computational Linguistics.

Sara Tonelli, Giuseppe Riccardi, Rashmi Prasad, and Aravind Joshi. 2010. Annotation of discourse relations for conversational spoken dialogs. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. 2019. The penn discourse treebank 3.0 annotation manual. *Philadelphia, University of Pennsylvania*, 35:108.

An Yang and Sujian Li. 2018. SciDTB: Discourse dependency TreeBank for scientific abstracts. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 444–449, Melbourne, Australia. Association for Computational Linguistics.

Amir Zeldes. 2017. The GUM Corpus: Creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51(3):581–612.

Deniz Zeyrek, Amália Mendes, and Murathan Kurfalı. 2018. Multilingual extension of PDTB-style annotation: The case of TED multilingual discourse bank. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Deniz Zeyrek Bozşahin, Ayışığı B. Sevdik Çallı, and Ruket Çakıcı. 2013. Turkish discourse bank: Porting a discourse annotation style to a morphologically rich language. *Dialogue and Discourse*, page 174–184.

Yuping Zhou, Jill Lu, Jennifer Zhang, and Nianwen Xue. 2014. Chinese discourse treebank 0.5. *LDC*.