

Genres in the Prague Discourse Treebank

Lucie Poláková, Pavlína Jínová, Jiří Mírovský

Charles University in Prague, Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

Malostranské nám. 25, 118 00 Prague 1, Czech Republic

{polakova,jinova,mirovsky}@ufal.mff.cuni.cz

Abstract

We present the project of classification of Prague Discourse Treebank documents (Czech journalistic texts) for their genres. Our main interest lies in opening the possibility to observe how text coherence is realized in different types (in the genre sense) of language data and, in the future, in exploring the ways of using genres as a feature for multi-sentence-level language technologies. In the paper, we first describe the motivation and the concept of the genre annotation, and briefly introduce the Prague Discourse Treebank. Then, we elaborate on the process of manual annotation of genres in the treebank, from the annotators' manual work to post-annotation checks and to the inter-annotator agreement measurements. The annotated genres are subsequently analyzed together with discourse relations (already annotated in the treebank) – we present distributions of the annotated genres and results of studying distinctions of distributions of discourse relations across the individual genres.

Keywords: discourse, genre classification, annotation

1. Introduction

Annotation of phenomena going beyond the sentence boundary is nowadays a well established field of corpus linguistics. There is an increasing number of corpora containing some type of discourse information (cf. for example the "family" of corpora in PDTB style annotated for discourse relations (Prasad et al., 2008, Oza et al., 2009, Al-Saif and Markert, 2010, Zhou and Xue, 2012, Zeyrek et al., 2010) or corpora containing coreference relations (e. g. OntoNotes (Pradhan et al., 2007), GNOME (Poesio, 2004), ARRAU (Poesio and Artstein, 2008)). Among other topics in this field, genre distinction is considered (cf. e. g. Webber, 2009) an important clue for experiments in NLP, mostly for those working with larger text units (anaphora resolution, text topics and salience, discourse processing, sentiment analysis etc.). Moreover, genre distinction is increasingly included also in the description of grammar of various languages as its inherent part (this tendency is most visible in various types of construction grammar, cf. e.g. Fried, Östman, 2004).

During the annotation of the Prague Discourse Treebank 1.0 (PDiT¹, Poláková et al., 2012), we experienced a considerable diversity of the data. Although the whole corpus consists of journalistic texts, it contains in fact texts ranging from TV programs to cultural reviews and also some number of unrelated texts in one document like short news collections (compare the text samples below²).

The manual classification of PDiT texts according to their genre or text style, described in this paper and newly included in the Prague Dependency Treebank 3.0

(PDT 3.0³, Bejček et al., 2013), should serve the following purposes: i) to exclude short and incoherent texts from training sets for modeling any type of coherence, ii) to cluster the data more efficiently for different subtasks, and iii) to obtain gold data for automatic genre/text type clustering.

This paper first describes the concept, process (Section 3) and evaluation (Section 4) of annotation of genres in PDiT. Further, some basic analyses of distributions of discourse relations in individual genres are presented as a first step in the direction of thinking about the role of the genre category in coherence analysis and modeling (Section 5).

2. Related Work

The main inspiration for genre distinction in the PDiT was, apart from our original purposes mentioned above, a similar work by B. Webber (2009), who tried to identify genres in the Wall Street Journal texts (Penn Treebank 2.0 data, PTB, Marcus et al., 1995) for different NLP purposes. In her paper, she claims that data annotation so far was primarily intended to serve for single-sentence tools for automated language analysis. We share her view that "...ignoring this variety [of corpus texts] may actually hinder the development of robust language technology for analysing and/or generating multi-sentence text. As such, it is worth considering genre in the PTB, since doing so can allow texts from different genres to be weighted differently when tools are being developed." (Webber, 2009, 675). Webber, inspired by RST-treebank founders' (Carlson et al., 2002) observations about the WSJ data, has introduced four major genre categories in the PTB documents: essays, summaries, letters and news. These categories show differences in their volume in the

1 <http://hdl.handle.net/11858/00-097C-0000-0008-E130-A>

2 or, more precisely, the English translations of the original Czech texts

3 <http://hdl.handle.net/11858/00-097C-0000-0023-1AAF-3>

| A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 |
|-----------------|----------------|-----------------|----------------|----------------|----------------|----------------|----------------|
| 428 news | 147 news | 118 description | 179 news | 157 news | 143 news | 165 news | 185 news |
| 124 description | 50 comment | 59 comment | 43 sport | 40 sport | 48 sport | 49 sport | 53 sport |
| 113 sport | 36 sport | 41 sport | 26 description | 29 caption | 46 essay | 25 comment | 28 comment |
| 68 essay | 22 description | 35 essay | 22 review | 28 description | 34 caption | 23 description | 18 caption |
| 60 other | 20 caption | 28 news | 17 comment | 23 comment | 22 comment | 23 caption | 14 review |
| 53 caption | 19 essay | 16 review | 12 caption | 10 review | 19 review | 13 review | 12 description |
| 38 review | 16 invitation | 14 other | 11 essay | 10 other | 14 description | 12 essay | 11 essay |

Table 1: Distributions of genres for the eight individual annotators (note that annotator A1 annotated more data than the others). Three frequent genres (news, description and sport) are highlighted by different colours for easier comparison. The slightly highlighted entries at the bottom of columns A1, A3, A5 and A6 mark misspelling errors.

treebank (news genre constitutes cca 88% of the PTB) and, more importantly, significant differences in the distributions of discourse connectives and other discourse-related phenomena.

With the decision to annotate genres on the Czech data of PDiT, we believe to make the rich multilayer annotations of the treebank even more valuable, and its outcome comparable to the previous attempts in this field.

3. Annotation

3.1. Data in Question – PDiT 1.0

We annotated genres on the data of the Prague Discourse Treebank 1.0 (PDiT, Poláková et al., 2012, 2013), which is an annotation extension of the Prague Dependency Treebank 2.5⁴ (PDT, Bejček et al., 2012)⁵. PDT is a corpus of almost 50 thousand sentences in 3,165 documents of Czech journalistic texts, the data originate from two big Czech daily newspapers (Mladá Fronta, Lidové Noviny), and one business weekly (Českomoravský profit). The corpus contains documents in the length of 1 - 231 sentences, with an average length of 15.6 sentences.

PDT was annotated manually on several layers of language description. The topmost (tectogrammatical) layer contains dependency trees capturing the deep syntax of the sentences, including annotation of pronominal textual coreference, multi-word expressions and information structure. In addition to the PDT annotation, PDiT contains extended annotation of coreference

relations, annotation of bridging anaphora, and explicitly realized discourse relations⁶.

For demonstration of data diversity in PD(i)T, three examples of different genres of the corpus documents are shown below: sport news (Example 1), a collection of unrelated texts (Example 2), and a photo caption (Example 3).

(1) *Jagr scores again*

New York -

The Czech hockey player Jaromir Jagr scored his fourteenth goal of the NHL season and so decided the result of the match Pittsburgh – Quebec (5:4). The final third was extremely dramatic, six goals were scored, and it was Jagr who had the last word and decided the match just 22 seconds after Nolan from Quebec leveled the score at 4:4. After being absent for four matches because of flu, Martin Straka joined the match and scored a goal.

In Miami, Florida was defeated by New York Rangers 3:5.

In the state of 3:3 in the final third, Karpovcev broke the tie when the puck from his stick ended up after a bounce in the opponent's net. The final 5:3 victory of the Rangers was decided by Olczyk.

(2) *Briefly*

Yesterday, the proposals of the British Prime Minister J. Major and his Irish partner J. Burton on the future organization of Northern Ireland received the support of the British government. The document will be a point of discussions of constitutional Northern Irish political parties.

The two main goals of the Czech foreign policy are the membership in the European Union and in NATO, said yesterday the Czech Minister of Foreign Affairs Josef Zieleniec in the Committee of Foreign Affairs and Foreign Trade Chamber

⁴ <http://hdl.handle.net/11858/00-097C-0000-0006-DB11-8>

⁵ All versions of PDT and the PDiT contain annotations over the same data.

⁶ i.e. discourse relations signaled by an overtly present discourse connective

of Deputies of the Parliament of Canada.

Another armed conflict between the army and the rebel organization Unita, which occurred at north Angola city of Uige, broke the agreement on ceasefire.

The Iraqi government keeps in "appalling" extent and "without any signs of improvement" trampling on human rights, says UN special reporter for Iraq, Max van der Stoep in his report, which was published yesterday at the Geneva UN headquarters.

So far, it cannot be said when new Polish government would be formed, said the coalition candidate for the seat of Prime Minister Marshal of the Sejm J. Oleksy after a meeting of representatives of Polish government coalition, the Polish People's Party and the Democratic Left Alliance.

(3) Dzhokhar Dudayev cannot deny being a former general of the Soviet Strategic Air, saluting perfectly at the festive parade organized on the occasion of the third anniversary of the declaration of independence of Chechnya from Russia. Photo Reuter

3.2. Annotation Scheme

In accordance with Carlsons et al's (2002) and Webber's (2009) observations about the genres in WSJ texts, we found out that the newspaper texts of PDiT can be classified to very similar subsets. In our project, the term "genre" is used in the same quite intuitive way, with respect to our data - in both PTB and PDiT the task is to further subcategorize journalistic style. We encouraged each of our discourse annotators to propose a genre specification scheme based on their previous experience with the corpus texts, then we merged their proposals into a finite one, which in the end very well corresponds to the main genre categories introduced for WSJ by Webber and so justifies our annotators' proposals. Finally we created a finer taxonomy of 20 categories in three main classes: monological genres, dialogical genres and other, marginal genres. Monological genres are the most varied class, it consists of the following categories: *critical review*, *invitation*, *letters from readers*, *advice column*, *cultural programme*, *film/TV program plot description*, *sports news*, *comment*, *news report*, *reflection essay*, *overview*, *description*, *weather forecast* and *readers' survey + its results*. The class of dialogical genres is divided into two categories, *topic interview* (emphasized is a given topic/multiple topics) and *person interview* (emphasized is the person being interviewed himself/herself). Other, marginal genres are *collections* (sets of unrelated texts, see Example 2), photo, table and chart *captions*, *metatexts* (metatext newspaper information occurred by mistake during corpus compilation) and there is an *other* category for occasional rare genres.

Attempting to classify texts according to their stylistic properties, we were naturally looking for criteria that would most fit the subset of journalistic stylistic forms we encounter in our data. Our taxonomy of genres therefore combines formal, content-based and pragmatic criteria: a (reader's) letter has a clearly distinguishable form as its main genre feature, while a weather forecast has a clearly delimited content.

| genre/feature | formal | | | content | | pragmatic |
|---------------|------------------------|------------------------------------|-------------------------------|--------------------------------------|---|---|
| | long (+)/ short (-) | specific document structure* | dialog (+)/ monolog (-) | specific content domain (+) | actual event (+)/ non- actual (-) | objective (+) /subjective (-) perspective |
| essay | + | - | - | - | n | - |
| comment | - | - | - | - | n | - |
| news | n | - | - | - | + | + |
| description | n | - | - | - | - | + |
| topic_interv | n | + | + | + | n | - |
| collection | n | + | - | - | n | + |
| person_interv | n | + | + | - | n | - |
| review | n | - | - | + | + | - |
| sport | n | - | - | + | + | + |
| letter | n | + | - | - | n | - |
| advice | n | + | - | - | n | n |
| survey | n | + | - | - | n | n |
| invitation | - | + | - | + | + | n |
| overview | n | + | - | - | n | + |
| caption | - | + | - | - | n | n |
| plot | n | - | - | + | n | + |
| program | n | + | - | + | + | + |
| weather | n | - | - | + | + | + |
| other | n | n | n | n | n | n |

* the feature *specific document structure* alone is in some cases sufficient enough for genre identification

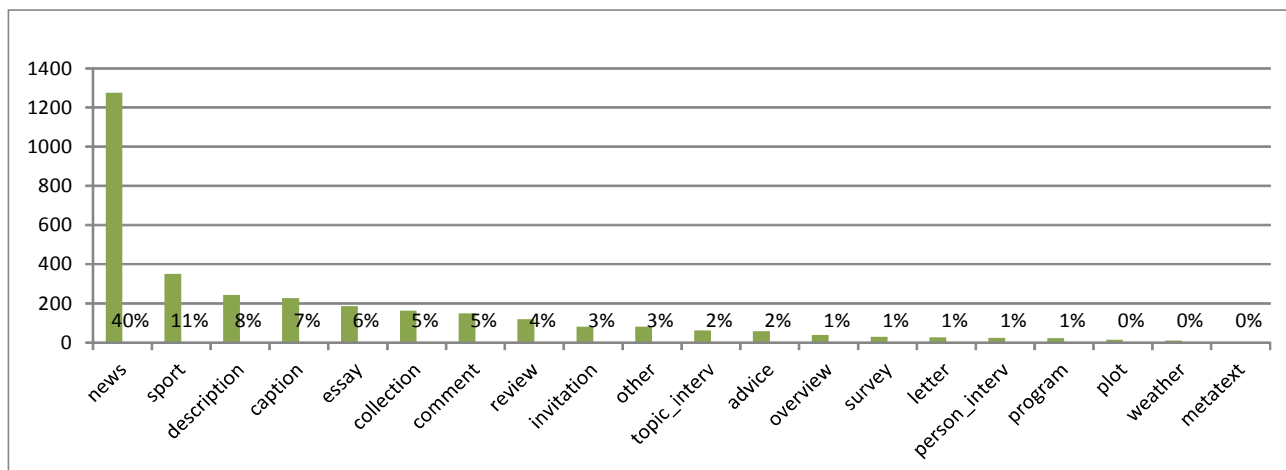
Table 2: List of distinctive features for the genres. The value "n" indicates that both "+" and "-" are possible values in the given attribute

Further, the characteristics of a news article in terms of its main genre features would be 1, **objective** (pragmatic feature) 2, **message** (formal feature) 3, about a **current event** (content feature), whereas an essay would be defined as 1, **subjective** (or evaluating, pragmatic feature) 2, **longer** (formal feature) text about a 3, **current event** or a **general state** (content feature). In this way, some of the features (or even their opposite values) became highly relevant for genre identification in our sense, and other do not play any role at all. For a detailed characteristics of genre features see Table 2, the distinctive features are highlighted in color.

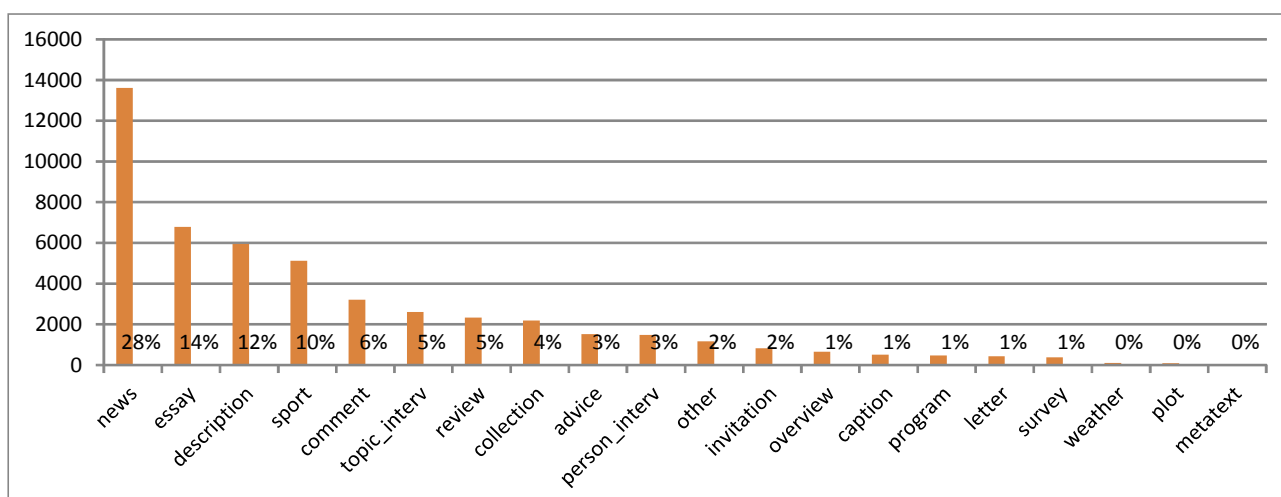
In the PDiT data notation, the genre information is represented by a document-level attribute "genre" of the tectogramatical document (it is neither an attribute of a tree nor of a node). For corpus querying purposes, however, this attribute has been subsequently mapped onto every t-tree root node. In this way, a search for any type of previously annotated linguistic information in combination with genre is enabled, e.g. "all headings (discourse-level attribute) of sports articles".

3.3. Annotation Procedure

The **automatic preannotation** used information from the manual annotation of discourse relations, where the annotators had marked corpus documents consisting of a set of short unrelated texts possibly of different genres (these were preannotated as *collections*) and also sentences representing photo, chart or table captions. In this way, 56 documents were preannotated as *captions* and 149 documents as *collections*.



Graph 1: Distribution of genres in PDiT documents



Graph 2: Distribution of genres in PDT sentences; in contrast with Graph 1, it shows higher uniformity of genres distribution in the data, reflecting different sizes of PDiT documents.

Annotation of the remaining documents was performed by eight annotators, Czech native speakers mostly with linguistic background. To keep the task as simple as possible, we only assigned one label to each document. In case of texts combining features of more than one genre (e.g. sports news with an incorporated interview with an athlete), annotators were instructed to mark the prevailing genre. 2/10 of the corpus (development test data and evaluation test data of PDiT) were annotated in parallel by two annotators.⁷ For most of the annotators (all but A1), it was also approx. 2/10 of their data. Discrepancies were then solved by an arbiter. In case of a substantial disagreement, the problematic genres were checked by the arbiter in all data annotated by the annotators in question (not only the parallel part).

4. Inter-Annotator Agreement

On the data annotated in parallel by two annotators (different pairs of annotators for different parts of

⁷ Every pair of annotators annotated a different part of the development and evaluation test data in parallel.

the data annotated in parallel), we measured the inter-annotator agreement. Table 3 shows the results. The first column indicates the annotators' pair, the middle column shows a simple agreement ratio between the annotators, and the right column contains the Cohen's κ .

| Annotators | Agreement (%) | Cohen's kappa |
|--------------|---------------|---------------|
| A1 vs. A2 | 58 | 0,47 |
| A3 vs. A4 | 33 | 0,26 |
| A5 vs. A6 | 75 | 0,69 |
| A7 vs. A8 | 80 | 0,73 |
| total | 61 | 0,52 |

Table 3: The inter-annotator agreement on data annotated in parallel

The disagreement between one pair of the annotators was surprisingly high. Distributions of genres annotated by the individual annotators helped reveal the most problematic cases (and the misspell errors; see Table 1). An arbiter

then checked and corrected the problematic genres in the annotation of the annotator whose understanding of the distinction of the genres was not correct (as indicated by a different distribution of genres). For example, most of the disagreements between annotators A3 and A4 were disagreements in genres *news* and *description*. The distributions revealed that annotator A3 chose *description* four times more often than *news*. For all the other annotators, the ratio was approx. the opposite one.

5. First Observations

For accessing this new type of information in annotated data for the first time, we performed a series of basic analyses of the data. First, we observed the overall final distribution of genres in PDiT (after the preannotation, manual annotation and corrections made by the arbiters), second, the frequency ("density") of annotated discourse relations (and connectives) in individual genres, and third, distributions of discourse senses in individual genres, where the similarity/difference between pairs of genres was measured using the relative entropy (Kullback–Leibler divergence).

5.1. Genre composition of the PD(i)T

Graph 1 and Graph 2 show distributions of genres in the documents and sentences of PDiT, respectively. At first sight we can see that in the sentences, the genres are distributed more uniformly. The differences between the graphs are given by the tendency of some genres to form longer (or shorter) documents, which is not surprising. For example, *news* is a genre of 40% of documents but at the same time of only 28% of all sentences in the corpus, which means that *news* documents tend to be shorter than an average document. The situation is opposite with *essay* and *description* – these two genres tend to appear in longer documents. *Sport* documents seem to be of average length, as in both graphs they form about the same part of the treebank (11% and 10%, respectively). Genres *program*, *caption* and *metatext* (and maybe also *collection* and *weather*) we consider unsuitable as training data for most NLP tasks (to exclude them from the data for this purpose was one of the goals we had in mind when we thought about the annotation of genres); from the graphs we can see that they form 8% of documents (13% if we count also *weather* and *collection*) but only 2% of sentences (6%, resp.), as they – maybe with the exception of *collection* – are usually quite short.

5.2. Frequency of discourse relations in genres

One aspect of the analysis worth looking into is the frequency of discourse relations, i.e. explicit discourse connectives across individual genres. By measuring the ratio between the number of sentences and the number of discourse relations in each genre (see Table 4) we made the observation that in the PDT journalistic data, explicit connectives are most frequently used in genres with a high degree of subjectivity, i.e. where opinions, desires, evaluations, beliefs etc. are expressed.

| genre | #sentences/ #relations | #sentences/ #inter-sent. relations | #sentences/ #intra-sent. relations |
|---------------|---------------------------|--|--|
| letter | 1.6 | 5.8 | 2.1 |
| essay | 1.8 | 5.6 | 2.7 |
| comment | 1.9 | 5.4 | 2.8 |
| person interv | 1.9 | 6.8 | 2.7 |
| sport | 2.0 | 9.3 | 2.6 |
| advice | 2.0 | 5.9 | 3.1 |
| topic interv | 2.1 | 6.4 | 3.1 |
| review | 2.2 | 7.1 | 3.1 |
| other | 2.2 | 9.5 | 2.9 |
| description | 2.6 | 7.9 | 3.8 |
| survey | 2.7 | 10.6 | 3.6 |
| news | 3.0 | 9.7 | 4.4 |
| invitation | 3.8 | 12.4 | 5.4 |
| collection | 4.3 | 17.5 | 5.7 |
| plot | 4.8 | 33.3 | 5.6 |
| overview | 4.9 | 22.1 | 6.4 |
| weather | 8.1 | 56.5 | 9.4 |
| caption | 13.3 | 101.4 | 15.4 |
| program | 18.3 | 53 | 28.1 |

Table 4: Ratio of number of sentences and number of discourse relations in each genre, also separately for inter-sentential and intra-sentential relations

With the exception of *sport*, the first eight positions are represented by genres in which a certain degree of subjectivity plays (or at least may play) an important role, while the "objective" genres gathered consistently lower in the connective frequency scale. In our opinion, *sport* has a relatively high position mainly due to the frequent reference to rapidly changing situations. On the other hand, *program* or *caption* are typical in containing only a minimum of connectives since they are either very short (in the case of *caption*) or they are often represented by verbless phrases only (both genres in view)⁸.

5.3. Sense (discourse type) distributions

We measured differences between distributions of discourse senses in pairs of genres using the relative entropy (Kullback–Leibler divergence), see Table 5 for a list of twelve closest and twelve most distant genres. We only considered genres with at least 500 occurrences of discourse relations and for each such genre, only 500 randomly selected discourse relations were counted. The statistical significance of the pair-wise differences between the genres was tested by the likelihood ratio test for multinomial distributions with low counts at some categories (Dunning, 1993). Differences in all pairs of discourse types distributions in genres were statistically significant with the confidence 0.05; differences between some closest pairs (measured by the relative entropy) were not significant with the confidence 0.01 (*comment* and *essay*, *personal interview* and *topic interview*, *personal interview* and *collection*), a few others were on the borderline.

⁸ The discourse annotation in PDiT only concerns verb containing structures as its units.

| genre 1 | genre 2 | KL-distance |
|---------------|---------------|-------------|
| comment | essay | 0.039 |
| person_interv | topic_interv | 0.039 |
| collection | person_interv | 0.040 |
| news | topic_interv | 0.047 |
| essay | person_interv | 0.048 |
| news | person_interv | 0.048 |
| collection | essay | 0.049 |
| collection | comment | 0.050 |
| collection | news | 0.053 |
| comment | topic_interv | 0.056 |
| comment | review | 0.057 |
| advice | topic_interv | 0.058 |
| ... | | |
| news | review | 0.101 |
| advice | description | 0.105 |
| description | person_interv | 0.106 |
| other | sport | 0.109 |
| advice | news | 0.110 |
| advice | collection | 0.110 |
| review | topic_interv | 0.113 |
| other | topic_interv | 0.119 |
| description | sport | 0.120 |
| advice | other | 0.122 |
| advice | review | 0.139 |
| advice | sport | 0.154 |

Table 5: Relative entropy (Kullback–Leibler divergence) for the closest and the most distant pairs of genres

Both inter-annotator agreement figures and the divergencies of discourse relation distributions across genres show that a few particular genre pairs are difficult to draw a line in between. This very well mirrors the smaller validity of their distinguishing features: the difference between e.g. *comment* and *essay* is basically in their length (and with it in the depth of reflection on the given topic). Such similarity suggests merging of these two categories into one larger group of subjective newspaper writing. For the purposes of data clustering for different NLP tasks, this step would create a larger set of homogenous data, so it is a desirable one. The same concerns the dialogical genres.

The other end of the scale shows significant differences between some of the "small" genres. This confirms our intuitive assumption that certain genres have a very specific structure. For instance, *advice* to readers shows the highest number of conditions (i.e. explicit conditional connectives): 13.8% to 7.4% in average distributions of senses. *Description* (e.g. of some company's business activities, of a certain phenomenon in society etc.) and *sports* news seem to represent a dichotomy of the most static and the most dynamic genres with highest numbers of oppositions and temporal asynchrony for *sports* and lower than average for *descriptions*. *Advice* on one hand vs. *news*, *collections* (of news) and *sport* on the other have then clearly different communicative functions: informative vs. instructive.

Contrary to our assumption that *news* and *description*, both being "big" genre categories, would prove quite alike since they have been a major issue in IAA disagreements,

their difference in discourse relations distribution is statistically significant (with relative entropy 0.069).

We are aware of the strong limits of our observations given by the type, size and occasional annotation sparsity of our data. We also do not make any claims about our genre categorization being the only possible one. We believe, though, that we have conducted valid experiments relevant for further (both linguistic and automated) work with the PD(i)T data.

6. Conclusion

From the inter-annotator agreement point of view, the genre annotation proved to be quite a challenging task. However, the subsequent analysis showed that most discrepancies are in genres that are close to each other and hard to draw a clear line in between. Therefore, we believe that for the purposes stated in the introduction, the quality of the genre annotation is good enough.

We have further shown that differences in distributions of discourse relations in genres in our data are mostly statistically significant. This is the first step in exploring the different ways coherence is established in different types of language data.

Last but not least, we can confirm, on a typologically different language than English, previous observations that genre is usable as a feature for text processing language technologies.

The genre annotation presented in this paper was published in 2013 as a part of the Prague Dependency Treebank, version 3.0, under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License and it is available to download from the LINDAT-Clarín repository⁹. It can be also easily backported to the published version of PDiT (1.0) and the previous versions of PDT (2.0, 2.5).

7. Acknowledgements

We gratefully acknowledge support from the Grant Agency of the Czech Republic (project n. P406/12/0658) and the SVV of the Charles University (project n. 260 104). This work has been using language resources developed, stored and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2010013). We also thank Jan Hajič Jr. for significant help with statistical significance tests.

8. References

- Al-Saif, A., Markert, K. (2010). The Leeds Arabic Discourse Treebank: Annotating discourse connectives for Arabic. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, Valletta, Malta, pp. 2046–2053.

⁹ <https://lindat.mff.cuni.cz/repository/xmlui/handle/11858/00-097C-0000-0023-1AAF-3>

- Bejček, E., Panevová, J., Popelka, J. et al. (2012). Prague Dependency Treebank 2.5 – a revisited version of PDT 2.0. In *Proceedings of the 24th International Conference on Computational Linguistics (Coling 2012)*, Mumbai, India, pp. 231–246.
- Bejček, E., Hajičová, E., Hajič, J., et al. (2013). *Prague Dependency Treebank 3.0*. Data/software, Univerzita Karlova v Praze, MFF, ÚFAL, Prague, Czech republic, <http://ufal.mff.cuni.cz/pdt3.0/>, Dec 2013.
- Carlson, L., Marcu, D. and Okurowski, M. E. (2002). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, Aalborg, Denmark.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. In: *Computational Linguistics - Special issue on using large corpora: I*, Volume 19, Issue 1, MIT Press Cambridge, MA, USA, March 1993, pp. 61–74.
- Marcus, M. P. et al. (1995): *Treebank 2*. Linguistic Data Consortium, CAT: LDC95T7, ISBN 1-58563-054-3. (Penn TreeBank 2.0).
- Mirjam, F., Östman, J.-O. (2004). Construction Grammar: a thumbnail sketch. In M. Fried & J.-O. Östman (Eds.), *Construction Grammar in a cross-language perspective*. Amsterdam & Philadelphia: John Benjamins, pp. 11–86.
- Oza, U., Prasad, R., Kolachina, S. et al. (2009). The Hindi Discourse Relation Bank. In *Proceedings of Linguistic Annotation Workshop*, pp.158–161.
- Poesio, M. (2004). The MATE/GNOME Proposals for Anaphoric Annotation, Revisited. In *Proceedings of The 5th SIGdial Workshop on Discourse and Dialogue*, Boston.
- Poesio, M., Artstein, R. (2008). Anaphoric annotation in the ARRAU corpus. In *Proceedings of LREC 2008*, Marrakech.
- Poláková, L., Mírovský, J., Nedoluzhko A. et al. (2013). Introducing the Prague Discourse Treebank 1.0. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, Nagoya, Japan, ISBN 978-4-9907348-0-0, pp. 91–99.
- Poláková, L., Jínová, P., Zikánová, Š. et al. (2012). *Prague Discourse Treebank 1.0*. Data/software, ÚFAL MFF UK, Prague, Czech Republic, <http://ufal.mff.cuni.cz/pdit/>, Nov. 2012.
- Pradhan, S., Hovy, E., Marcus, M. et al. (2007). Ontonotes: A unified relational semantic representation. In *Proceedings of the International Conference on Semantic Computing*, Washington DC.
- Prasad, R., Dinesh, N., Lee, A. et al. (2008). The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco, pp. 2961–2968.
- Webber, B. (2009). Genre distinctions for discourse in the Penn TreeBank. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*. Association for Computational Linguistics.
- Zeyrek, D., Demirşahin, I., Sevdik-Çalli, A. et al. (2010). The Annotation Scheme of the Turkish Discourse Bank and an Evaluation of Inconsistent Annotations. In *Proceedings of the Fourth Linguistic Annotation Workshop*, Uppsala, Sweden, pp. 282–289.
- Zhou, Y., Xue, N. (2012). PDTB-style Discourse Annotation of Chinese Text. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*. Jeju, Republic of Korea, pp. 69–77.