A Multi-Dimensional Analysis of RST Discourse Relations in Eight Genres

This paper examines genre variation from the viewpoint of discourse relations in the framework of Rhetorical Structure Theory (RST, Mann & Thompson 1988). Discourse relations, such as 'cause' or 'condition' in Figure 1, are signaled using a range of linguistic means (Das & Taboada 2017), including discourse markers, syntactic devices (e.g. relatives), semantics (e.g. co-referentiality) and genre conventions (e.g. use of headings).
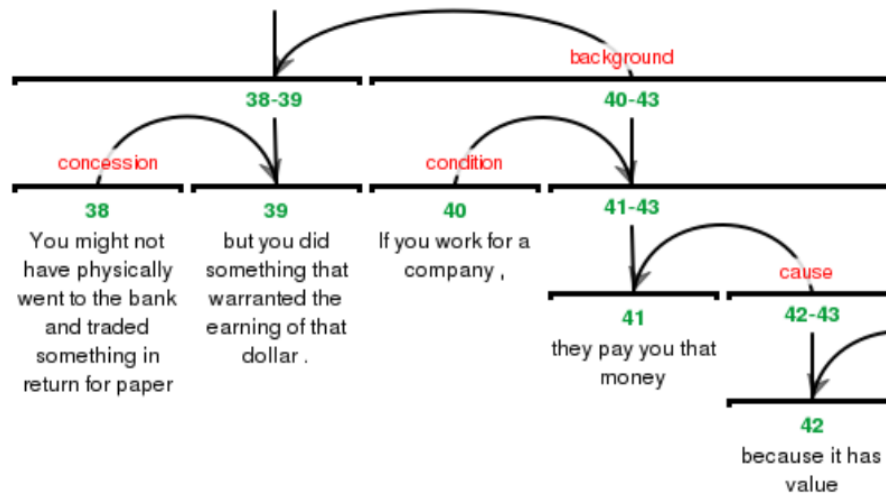

Figure 1. Fragment from a discourse relation analysis.

While Multi-Dimensional analyses (Biber 2009) have shown the importance of lexical, morphological and syntactic cues across genres, genre studies of discourse relations remain few (see Taboada & Lavid 2003, Gruber & Huemer 2008). Using the Georgetown University Multilayer corpus (GUM, Zeldes 2017), we compare data covering 85,000 tokens/7,400 instances of 20 relations from eight text-types taken from the Web: news, interviews, how-to guides, travel guides, reddit forums, and academic, biographical, and fiction writing.

We discuss and integrate two kinds of results into a Multi-Dimensional analysis: on one hand, genres vary widely in proportions and arrangements of discourse relations. On the other, ways in which the same relations are signaled depend strongly on genre. For example, we can consider the contrastive 'contrast', 'antithesis' and 'concession' in Figure 2.
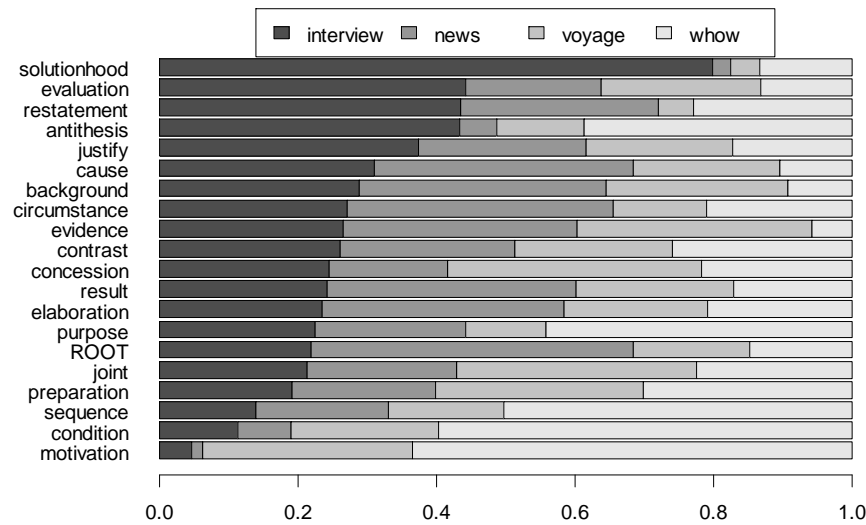
Figure 2. Relation proportions in four genres.

While neutral 'contrast' is balanced, 'antithesis' (presenting an alternative to favor the first option), is used most in interviews (often polemically), and in how-to guides ("whow" in Figure 2) to warn against not following instructions, but less in news and travel guides. A *concession*, which concedes value in an opposing idea, is common in travel guides ("voyage"), but less in news, as viable alternatives are emphasized in instructional texts, but not in news narrative. Starting from these distributions, we use machine learning methods to explore recurring words and annotations signaling the same relations in different genres.

References

Biber, Douglas (2009), Multi-Dimensional Approaches. In: Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics. An International Handbook*. Vol. 2. Berlin: Mouton de Gruyter, 822–855.

Das, Debopam & Maite Taboada (2017), Signalling of Coherence Relations in Discourse, Beyond Discourse Markers. *Discourse Processes*, online first, 1-29.

Gruber, Helmut & Birgit Huemer (2008), Two Views on Text Structure: Using Rhetorical Structure Theory and Register & Genre Theory in Improving Students' Academic Writing. In: Nina Nørgaard (ed.), *Systemic Functional Linguistics in Use*. (Odense Working Papers in Language and Communication 29.) Odense: University of Southern Denmark, 332–364.

Mann, William C. & Sandra A. Thompson (1988), Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8(3), 243–281.

Taboada, Maite & Julia Lavid (2003), Rhetorical and Thematic Patterns in Scheduling Dialogues: A Generic Characterization. *Functions of Language* 10(2), 147–179.

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