

and the compositionality of meaning.

Part II: Paraphrase Typology and Paraphrase Identification

The three articles presented in this part of the thesis tell a coherent story of how the theoretical concepts of Paraphrase Typology (PT) research can be validated empirically and, at the same time, can be used to improve the evaluation, interpretation, and, indirectly, the performance of the automated Paraphrase Identification (PI) systems.

In the article “*WARP-Text: a Web-Based Tool for Annotating Relationships between Pairs of Texts.*” (Chapter 4), I describe the workings of a novel web-based annotation interface. The annotation interfaces that existed at the beginning of this thesis were not capable of performing a simultaneous annotation of multiple texts with fine-grained linguistic phenomena. WARP-Text fills this gap in the CL and NLP toolbox and creates new opportunities for researchers.

In the article “*ETPC - a paraphrase identification corpus annotated with extended paraphrase typology and negation*” (Chapter 5), I present the first PI corpus annotated with paraphrase types. I also propose a new extended typology for the paraphrasing relation, enriching the existing work in the area. I analyze the distribution of different linguistic phenomena in the corpus, and I identify general tendencies and potential biases in the data. This corpus-based study is the first large-scale empirical research on Paraphrase Typology within Paraphrase Identification. It contrasts with the pre-existing work in the field, in which researchers typically annotate a small number of examples. This article is also the first work on paraphrase typology that analyzes both positive examples (paraphrases) and negative examples (non-paraphrases). It contrasts with the pre-existing work in the field, which focuses only on the positive examples. The ETPC corpus makes further machine learning based studies on Paraphrase Typology possible.

In the article “*A Qualitative Evaluation Framework for Paraphrase Identification*” (Chapter 6), I perform multiple machine learning experiments on the ETPC corpus. I re-implement 11 different machine learning systems and create an “evaluation framework” - a software package that can quantify and compare the paraphrase types involved in the correct and incorrect prediction of each PI system. I empirically demonstrate that 1) the different paraphrase types are processed differently by the different state-of-the-art automated PI systems; and 2) some paraphrase types are easier or harder for all evaluated systems. Furthermore, I demonstrate that the “qualitative evaluation framework” provides much more information when comparing automated systems and facilitates error analysis.

Part III: A Joint Study of Meaning Relations

The two articles presented in this part of the thesis demonstrate that multiple textual meaning relations can co-exist in the same corpus and can be expressed using the same “atomic” linguistic phenomena.

In the article “*Annotating and analyzing the interactions between meaning relations*”, I present the first corpus to explicitly annotate the meaning relations of Paraphrasing, Textual Entailment, Contradiction, Semantic Similarity, and Textual Specificity. I propose a methodology for corpus creation and annotation that guarantees that all relations are presented with a sufficient frequency. I compare the reliability of the annotation and the inter-annotator agreement across all relations. Finally, I perform an empirical analysis of the frequency, correlation, and overlap between the different meaning relations.

In the article “*Decomposing and Comparing Meaning Relations: Paraphrasing, Textual Entailment, Contradiction, and Specificity*” I propose SHARel - a shared typology for Paraphrasing, Textual Entailment, Contradiction, Semantic Similarity, and Textual Specificity. I demonstrate that a single typology can successfully be applied to all textual meaning relations. I analyze the distribution of the types across all relations and I outline common tendencies and differences.

1.4 Thesis Outline

This thesis consists of a collection of seven papers, complemented by an introductory and a concluding chapter that provide the necessary context to make the thesis a coherent story. The seven papers are the following:

Part I: Similarity at the Level of Words and Phrases

1. Venelin Kovatchev, M. Antònia Martí, and Maria Salamó. 2016. Comparing Distributional Semantics Models for identifying groups of semantically related words. *Procesamiento del Lenguaje Natural* vol. 57, pp.: 109-116
2. M. Antònia Martí, Mariona Taulé, Venelin Kovatchev, and Maria Salamó. 2019. DISCover: DIStributional approach based on syntactic dependencies for discovering COnstructions. *Corpus Linguistics and Linguistic Theory*

Part II: Paraphrase Typology and Paraphrase Identification

3. Venelin Kovatchev, M. Antònia Martí, and Maria Salamó. 2018. WARP-Text: a Web-Based Tool for Annotating Relationships between Pairs of Texts. *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pp.: 132-136

4. Venelin Kovatchev, M. Antònia Martí, and Maria Salamó. 2018. ETPC - a paraphrase identification corpus annotated with extended paraphrase typology and negation. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, pp.: 1384-1392
5. Venelin Kovatchev, M. Antònia Martí, Maria Salamó, and Javier Beltran. 2019. A Qualitative Evaluation Framework for Paraphrase Identification. *Proceedings of the Twelfth Recent Advances in Natural Language Processing Conference*, pp.: 569-579

Part III: Paraphrasing, Textual Entailment, and Semantic Similarity

6. Darina Gold, Venelin Kovatchev, Torsten Zesch. 2019. Annotating and analyzing the interactions between meaning relations *Proceedings of the Thirteenth Language Annotation Workshop*, pp.: 26-36
7. Venelin Kovatchev, Darina Gold, M. Antònia Martí, Maria Salamó, and Torsten Zesch. 2020. Decomposing and Comparing Meaning Relations: Paraphrasing, Textual Entailment, Contradiction, and Specificity. To appear in *Proceedings of the Twelfth International Conference on Language Resources and Evaluation*, 2020

All seven papers have been accepted and published in peer reviewed journals or conference proceedings. They are co-authored by both my advisors, with one exception: Paper 6 was written during my research stay at the University of Duisburg-Essen and is co-authored with the Language Technology Group at that university. In all papers, except paper 2, I am listed as the first author. In paper 2, I was responsible for the evaluation section of the article and part of the experimental setup. In paper 6, the first two authors (Darina Gold and myself) contributed equally to the article and the names are in alphabetical order.

The papers reprinted here have been reformatted to make the typography of the thesis consistent, and all of the references and appendices have been integrated in a single bibliography and appendix section at the end. The thesis also includes some additional material, such as annotation guidelines, which were included in the original papers as external web links.