

VU Research Portal

Representative Resources for Perspective-Aware Natural Language Inference

van Son, Chantal Michelle

2024

DOI (link to publisher)
[10.5463/thesis.644](https://doi.org/10.5463/thesis.644)

document version
Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

van Son, C. M. (2024). *Representative Resources for Perspective-Aware Natural Language Inference*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam]. <https://doi.org/10.5463/thesis.644>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

CONTENTS

ACKNOWLEDGEMENTS	v
LIST OF FIGURES	xiv
LIST OF TABLES	xvi
LIST OF ABBREVIATIONS	xviii
LIST OF LANGUAGE RESOURCES	xix

INTRODUCTION	1
--------------	---

1 INTRODUCTION	3
1.1 Motivation	3
1.2 The complexities of natural online debate	5
1.3 Understanding propositions, perspectives, and inference	6
1.3.1 Propositions	7
1.3.2 Perspectives	7
1.3.3 Inference	8
1.4 The importance and limitations of language resources	9
1.5 Research questions	10
1.6 Structure of the thesis	11
1.7 Contributions	14
1.7.1 Publications	15
1.7.2 Data and software	19
1.8 Disclaimer	19

I UNDERSTANDING PROPOSITIONS THROUGH EVENT REPRESENTATIONS	21
--	----

2 INVESTIGATING INTEROPERABILITY: THE CASE OF EVENTS	27
2.1 Introduction	27
2.2 Structure of the chapter	28
2.3 Overview of event corpora	28
2.3.1 PropBank, OntoNotes & AMR	29
2.3.2 ACE & ERE	29
2.3.3 TimeML	30
2.3.4 Other annotation standards	31

CONTENTS

2.4	Defining interoperability	31
2.5	Document interoperability	32
2.5.1	Mapping document identifiers	32
2.5.2	Document intersections	33
2.6	Annotation interoperability: a small-scale analysis	34
2.6.1	Annotation guidelines	37
2.6.2	Data formats	38
2.6.3	Sentence and token identifiers	39
2.6.4	Approach to resolving structural interoperability	39
2.6.5	Type-based analysis	40
2.6.6	Annotation overlap	42
2.6.7	Consistency of annotations	44
2.7	Discussion	45
2.8	Conclusion	47
II	UNDERSTANDING PERSPECTIVES THROUGH SEMANTIC EFFECTS OF NEGATION	51
3	CAPTURING SEMANTIC SUBTYPES OF LEXICAL NEGATION	57
3.1	Introduction	57
3.2	Structure of the chapter	58
3.3	Defining lexical negation	58
3.3.1	Affixal negation	58
3.3.2	Regular antonyms	59
3.3.3	Semantic subtypes of affixal negation	60
3.3.4	Lexical negation in existing language resources	60
3.4	Building a negation dictionary	62
3.4.1	Antonyms from WordNet	62
3.4.2	Annotation tasks	63
3.4.3	Evaluation	64
3.5	Discussion	65
3.5.1	Effectiveness of Joshi’s typology	65
3.5.2	Inflections of verbal affixal negations	66
3.5.3	Opposition versus affix-base relations	67
3.5.4	Coverage and fallback strategies	67
3.6	Conclusion	68
4	EXPOSING PRAGMATIC INFERENCE OF CLAUSAL NEGATION	69
4.1	Introduction	69
4.2	Structure of the chapter	70
4.3	Related work	71
4.4	Dataset and task	72
4.4.1	Data access and interoperability	73

4.5	Regression task	74
4.5.1	Experimental set-up	74
4.5.2	Results	76
4.6	Classification task	80
4.6.1	Experimental set-up	80
4.6.2	Results	81
4.6.3	Error analysis	82
4.7	Discussion	85
4.8	Conclusion	87
 III PERSPECTIVE-AWARE NATURAL LANGUAGE INFERENCE		89
5	BUILDING THE PERSPECTIVE-AWARE NLI DATASET	99
5.1	Introduction	99
5.2	Structure of the chapter	101
5.3	The Vaccination Corpus	101
5.4	Challenges in extracting pairs from natural online debate	103
5.5	Challenge 1: Extracting statements	105
5.5.1	Handling dependent clauses	105
5.5.2	Excluding negation and discourse markers	106
5.6	Challenge 2: Pairing premises and hypotheses	107
5.6.1	Selection criteria	107
5.6.2	Intra-sentence matching	109
5.6.3	Inter-sentence matching	110
5.7	Challenge 3: Creating perspective annotation units	111
5.8	Crowd annotation	112
5.8.1	Task design	113
5.8.2	Quality control	115
5.8.3	Participants	116
5.9	Discussion	118
5.10	Conclusion	119
6	MEASURING AMBIGUITY: A CROWDTRUTH ANALYSIS	121
6.1	Introduction	121
6.2	Structure of the chapter	122
6.3	CrowdTruth framework	122
6.4	Annotation quality scores	124
6.4.1	Common confusions	124
6.5	Unit quality scores	125
6.5.1	Effect of relation type	126
6.5.2	Effect of additional sources	126
6.5.3	Effect of textual similarity	127
6.6	Linguistic phenomena	128

CONTENTS

6.6.1	Intra-sentence pairs	130
6.6.2	Inter-sentence pairs	135
6.7	Discussion	140
6.8	Conclusion	143
7	INVESTIGATING BIAS: A TRANSFORMER MODEL EVALUATION	145
7.1	Introduction	145
7.2	Structure of the chapter	146
7.3	The MultiNLI corpus	147
7.4	Preparing PANLI for experiments	147
7.4.1	PANLI-3 and PANLI-4	147
7.4.2	Source integration	149
7.4.3	Data splitting strategy	150
7.5	DeBERTa: a Transformer-based language model	150
7.5.1	The DeBERTa model	151
7.5.2	Implementation details	152
7.5.3	Partial models	152
7.6	Methodological framework for error analysis	153
7.6.1	SHapley Additive exPlanations (SHAP)	154
7.6.2	Lexical types	155
7.6.3	Lexical relations	155
7.6.4	Event-selecting predicates	157
7.7	Main results	159
7.7.1	Full models	159
7.7.2	Partial models	159
7.8	Error analysis I: Pair types	160
7.8.1	Relation type	161
7.8.2	Source type	162
7.9	Error analysis II: Model confidence	163
7.9.1	Relation between human and model confidence	164
7.9.2	Relation between textual similarity and model confidence	165
7.10	Error analysis III: Explaining bias using SHAP	165
7.10.1	Hypothesis-only model	165
7.10.2	Premise-only model	168
7.11	Error analysis IV: Linguistic phenomena	170
7.11.1	Lexical types	170
7.11.2	Lexical relations	170
7.11.3	Event-selecting predicates	172
7.12	Error analysis V: PANLI-3 versus PANLI-4	173
7.13	Discussion	175
7.14	Conclusion	176

CONCLUSIONS	179
8 CONCLUSIONS	181
8.1 Observations on existing resources (RQ1)	181
8.1.1 Lack of interoperability	182
8.1.2 Limited size and class distribution	182
8.1.3 Lack of representativeness	183
8.1.4 Presence of bias	183
8.1.5 Absence of perspectives	183
8.2 Recommendations for building resources (RQ2)	184
8.2.1 Utilize real-world data	184
8.2.2 Minimize curation and post-processing	185
8.2.3 Examine and address bias	185
8.2.4 Integrate the notion of perspective	185
8.2.5 Acknowledge the nuances of meaning in natural language	186
8.2.6 Acknowledge the subjective nature of human interpretation	186
8.3 Contributions and limitations	186
8.3.1 Strengths of the PANLI dataset	187
8.3.2 Limitations of the PANLI dataset	187
8.4 Future research	188
8.4.1 Expanding and refining the PANLI dataset	188
8.4.2 Assessing the capabilities of various model architectures	189
8.4.3 Integrating ambiguity in model training and evaluation	190
8.5 Closing remarks	190
APPENDIX	191
A PANLI TASK DESIGN	193
A.1 Task summary	193
A.2 Instructions	194
A.3 Examples	195
B PANLI MODEL EVALUATION	197
B.1 Hyperparameters for fine-tuning DeBERTa	197
B.2 Confusion matrices	198
BIBLIOGRAPHY	201
ENGLISH SUMMARY	217
NEDERLANDSE SAMENVATTING	221