
CONTENTS

Abstract	ii
Acknowledgements	iii
List of Tables	xi
List of Figures	xvi
List of Publications	xix
Introduction	xxii
I Semantic Textual Similarity	1
1 Introduction to Semantic Textual Similarity	2
1.1 Datasets	8
1.1.1 English Datasets	9
1.1.2 Datasets on Other Languages	24
1.1.3 Datasets on Different Domains	33
1.2 Evaluation Metrics	36
1.3 Conclusion	40
2 Vector Aggregation based STS Methods	42
2.1 Related Work	44
2.1.1 Cosine Similarity on Average Vectors	45
2.1.2 Word Mover's Distance	45
2.1.3 Cosine Similarity Using Smooth Inverse Frequency	46

2.2	Improving State of the Art STS Methods	50
2.3	Portability to Other Languages	60
2.4	Portability to Other Domains	62
2.5	Conclusions	63
3	STS with Neural Sentence Encoders	66
3.1	Related Work	69
3.2	Exploring Sentence Encoders in English STS	74
3.3	Portability to Other Languages	77
3.4	Portability to Other Domains	80
3.5	Conclusions	82
4	Siamese Neural Networks for STS	84
4.1	Related Work	87
4.2	Exploring Siamese Neural Networks for STS	89
4.2.1	Impact of Transfer Learning	94
4.2.2	Impact of Data Augmentation	96
4.3	Portability to Other Languages	99
4.4	Portability to Other Domains	102
4.5	Conclusions	105
5	Adapting Transformers for STS	108
5.1	Related Work	111
5.2	Transformer Architecture for STS	114
5.3	Exploring Transformers in English STS	115
5.3.1	Impact of Transfer Learning	118

5.3.2	Impact of Data Augmentation	119
5.4	Portability to Other Languages	121
5.5	Potability to Other Domains	123
5.6	Recent Developments:	
	Siamese Transformers	126
5.7	Conclusions	128
II	Applications - Translation Memories	131
6	Introduction to Translation Memories	132
6.1	Related Work	136
6.2	Dataset	139
6.3	Evaluation	140
6.4	Conclusions	141
7	Sentence Encoders for Translation Memories	143
7.1	Motivation	146
7.2	Methodology	148
7.3	Results and Evaluation	151
7.4	Error Analysis	154
7.5	Conclusions	161
III	Applications - Translation Quality Estimation	163
8	Introduction to Translation Quality Estimation	164
8.1	Related Work	168

8.2	Datasets	172
8.2.1	Sentence-level QE	172
8.2.2	Word-level QE	175
8.3	Evaluation	177
8.4	Conclusion	178
9	TransQuest: STS Architectures for QE	180
9.1	Methodology	183
9.1.1	Experimental Setup	187
9.2	Results and Discussion	188
9.3	Further Fine-tuning	193
9.4	Error analysis	196
9.5	Conclusion	198
10	Extending TransQuest for Word-Level QE	201
10.1	Methodology	205
10.1.1	Experimental Setup	207
10.2	Results and Evaluation	207
10.3	Further Improvements	212
10.4	Conclusion	213
11	Multilingual Quality Estimation with TransQuest	216
11.1	Methodology	219
11.2	Results	220
11.2.1	Multilingual QE	221
11.2.2	Zero-shot QE	225

11.2.3 Few-shot QE	228
11.3 Conclusion	230
12 Conclusions	233
12.1 Research Questions	233
12.2 Achievements	235
12.3 Summary	240
Bibliography	241