
Contents

1	Introduction	1
1.1	Scientific Contributions	4
2	Background: Statistical Machine Translation and Computer-assisted Translation	7
2.1	Terminology	7
2.2	Statistical Machine Translation	8
2.2.1	Sentence Alignment	8
2.2.2	Word Alignment	9
2.2.3	Translation Models and Decoding	10
2.2.4	Additional Generative Models: Language and Reordering Model	11
2.2.5	Evaluation Metrics	12
2.2.6	Log-linear Model and Discriminative Parameter Training	13
2.3	Computer-assisted Translation	15
2.3.1	Translation Memory	15
2.3.2	MT Post-Editing	16
2.3.3	Interactive Translation	16
3	Detecting Cross-lingual Textual Entailment	17
3.1	CLTE Task	18
3.2	Related Work	21
3.2.1	The 2012 and 2013 SemEval CLTE for Content Synchronization Task	22
3.3	Modeling CLTE with SMT Features	22
3.3.1	Length-based Features	23
3.3.2	Cross-lingual Alignment Features	24
3.3.3	Monolingual Similarity Metrics	26
3.3.4	Meteor Features	27

3.3.5	Additional Meteor Alignment Features	28
3.4	Classification	29
3.5	Experimental Results	31
3.5.1	Baselines	32
3.5.2	Choosing a Classifier Setup	32
3.5.3	Individual Feature Results on Development Data	33
3.5.4	Results on SemEval12 CLTE Test Set	35
3.5.5	System Comparison	36
3.6	Discussion	37
3.6.1	Discarded Features	37
4	Quantifying Cross-lingual Similarity in a Noisy-Parallel Corpus	41
4.1	Corpus Construction	44
4.2	Annotation Procedure	46
4.2.1	Preliminary Experiment	47
4.2.2	Main Annotation	50
4.3	Adapting a System for Recognizing CLTE for Filtering	52
4.3.1	Binarizing the Annotation	52
4.3.2	Feature Engineering	53
4.4	Experimental Results	55
4.4.1	SMT System and Data	56
4.4.2	Baselines	57
4.4.3	Results and Discussion	58
5	Cross-lingual Fuzzy Match Retrieval for SMT	59
5.1	A Model for Biasing an SMT System with TM Fuzzy Matches	60
5.1.1	Related Work	62
5.2	Monolingual and Cross-lingual Fuzzy Match Retrieval	65
5.2.1	Locality Sensitive Hashing for Coarse Translation Retrieval	66
5.2.2	Fine-grained Selection of the Best Match	71
5.3	N-best Re-ranking with TM Matches	77
5.4	Experimental Results	78
5.4.1	Data Sets	78
5.4.2	Baseline SMT	80
5.4.3	Retrieval Parameters	81
5.4.4	Results	82
5.5	Analysis and Discussion	87

6	Online SMT Adaptation with User Feedback	91
6.1	Related Work	93
6.2	Online Adaptation for CAT	94
6.2.1	Constrained Search Alignment for Feedback Exploitation	96
6.2.2	Translation Model Adaptation: A Local Phrase Table	97
6.2.3	Language Model Adaptation: N-gram Cache Feature	99
6.3	Experimental Results	100
6.3.1	Data	100
6.3.2	Evaluation	102
6.3.3	Translation Model Adaptation	102
6.3.4	Language Model Adaptation	103
6.3.5	Results on Test Sets	104
6.4	Discussion	104
6.4.1	Example Output	105
6.4.2	Delayed Feedback	106
7	Summary of Contributions and Concluding Remarks	107
	Bibliography	111
	Acknowledgements	129