

# CONTENTS

<b>Acknowledgements</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>Preface</b>	<b>xiii</b>
<b>I Introduction and Goals</b>	<b>1</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Natural Language Processing . . . . .	3
1.2 Machine Translation . . . . .	3
1.2.1 MT Systems Taxonomy . . . . .	4
1.2.2 Rule-Based Systems . . . . .	5
1.2.3 Corpus-Based Systems . . . . .	6
1.3 Statistical Machine Translation . . . . .	7
1.4 Statistical Models for Machine Translation . . . . .	10
1.4.1 $n$ -gram Language Models . . . . .	11
1.4.2 Single-Word Alignment Models . . . . .	13
1.4.3 Multi-Word Alignment Models . . . . .	15
1.4.4 Log-Linear Models . . . . .	18
1.5 Parameter Estimation Techniques . . . . .	19
1.5.1 The Expectation-Maximisation Algorithm . . . . .	20
1.5.2 Generalised Iterative Scaling . . . . .	21
1.5.3 Minimum Error Rate Training . . . . .	21
1.6 Search Algorithms . . . . .	21
1.7 Alternative Techniques for Corpus-Based MT . . . . .	22
1.7.1 MT based on Stochastic Finite State Transducers . . . . .	22
1.7.2 MT based on synchronous context-free grammars . . . . .	23
1.8 Interactive Machine Translation . . . . .	24
1.9 Evaluation . . . . .	26
1.9.1 MT evaluation . . . . .	26
1.9.2 Word Alignment Evaluation . . . . .	27
1.9.3 IMT Evaluation . . . . .	27
1.10 Corpus . . . . .	28
1.10.1 EuTrans-I Corpus . . . . .	28
1.10.2 Europarl Corpus . . . . .	29
1.10.3 Hansards Corpus . . . . .	30

1.10.4	Xerox Corpus . . . . .	30
1.10.5	EU Corpus . . . . .	31
1.11	Summary . . . . .	31
<b>2</b>	<b>Scientific and Technologic Goals</b>	<b>33</b>
<b>II</b>	<b>Fully-Automatic Phrase-Based Statistical Machine Translation</b>	<b>37</b>
<b>3</b>	<b>Phrase-Based Statistical Machine Translation</b>	<b>39</b>
3.1	Introduction . . . . .	39
3.2	Standard Estimation of Phrase-Based Models . . . . .	39
3.2.1	Implementation Details . . . . .	42
3.3	Bisegmentation-based RF Estimation . . . . .	42
3.3.1	Complexity Issues . . . . .	43
3.3.2	Algorithm . . . . .	46
3.3.3	Implementation Details . . . . .	46
3.3.4	Possible Extensions and Applications of the Proposed Algorithms . . . . .	52
3.4	Phrase-based Model Estimation from Very Large Corpora . . . . .	55
3.4.1	Some Model Statistics . . . . .	55
3.4.2	Training Procedure . . . . .	55
3.5	Specific Phrase-Based Model Derivation . . . . .	58
3.5.1	Generative Process . . . . .	59
3.5.2	Model Derivation . . . . .	60
3.5.3	Log-Linear Model . . . . .	63
3.6	Summary . . . . .	64
<b>4</b>	<b>Search for Phrase-Based Statistical Machine Translation</b>	<b>65</b>
4.1	Introduction . . . . .	65
4.2	Branch-and-Bound Search for PB-SMT . . . . .	65
4.2.1	Dynamic Programming Algorithm . . . . .	67
4.2.2	Branch-and-Bound Search for PB-SMT . . . . .	72
4.2.3	Monotonic Search . . . . .	76
4.2.4	Stack Pruning and Multiple Stacks . . . . .	77
4.2.5	Breadth-First Search . . . . .	78
4.2.6	Generalised Multiple-Stack Algorithm for Best-First Search . . . . .	79
4.2.7	Generalised Multiple-Stack Algorithm for Breadth-First Search . . . . .	82
4.2.8	Additional Pruning Techniques . . . . .	83
4.2.9	Rest Score Estimation . . . . .	83
4.2.10	Generation of Word Graphs . . . . .	84
4.3	Efficient Decoding with Very Large Phrase-Based Models . . . . .	85
4.3.1	Cache Memory Architecture . . . . .	86
4.3.2	Selecting a Suitable Data Structure for Phrase Pairs . . . . .	87
4.4	Phrase-Level Alignment Generation . . . . .	90
4.4.1	Search Algorithm . . . . .	91

4.4.2	Smoothing Techniques . . . . .	94
4.4.3	A loglinear Approach to Phrase-to-Phrase Alignments . . . . .	97
4.5	Summary . . . . .	99
<b>5</b>	<b>PB-SMT Evaluation</b>	<b>101</b>
5.1	Phrase Model Estimation from Very Large Corpora . . . . .	101
5.2	Best- versus Breadth-First Search . . . . .	102
5.2.1	Assessment Criteria . . . . .	103
5.2.2	EuTrans-I Experiments . . . . .	104
5.2.3	Xerox Experiments . . . . .	105
5.2.4	Europarl Experiments . . . . .	106
5.3	Generalised Multiple-Stack Search . . . . .	106
5.3.1	Best-First Search Experiments . . . . .	106
5.3.2	Breadth-First Search Experiments . . . . .	110
5.4	Log-linear Model Performance . . . . .	113
5.4.1	Decoder Configuration . . . . .	113
5.4.2	Europarl Experiments . . . . .	114
5.4.3	Additional Experiments . . . . .	116
5.5	Bisegmentation-based RF Estimation . . . . .	117
5.6	Efficient Decoding with Large Phrase-Based Models . . . . .	119
5.7	Phrase-Level Alignment Generation . . . . .	120
5.7.1	Aligner Configuration . . . . .	121
5.7.2	Assessment Criteria . . . . .	121
5.7.3	Alignment Quality Results . . . . .	122
5.8	Summary . . . . .	124
<b>III</b>	<b>Interactive Phrase-Based Statistical Machine Translation</b>	<b>127</b>
<b>6</b>	<b>Interactive Phrase-Based Machine Translation</b>	<b>129</b>
6.1	Introduction . . . . .	129
6.2	IMT based on Partial Phrase-Based Alignments . . . . .	130
6.2.1	Search Algorithm . . . . .	131
6.2.2	Scoring Function . . . . .	133
6.3	IMT based on Stochastic Error-Correction Models . . . . .	134
6.3.1	PFSMs as Stochastic Error-Correction Models . . . . .	135
6.3.2	Alternative IMT Formalisation . . . . .	138
6.3.3	Generalisation . . . . .	143
6.4	Summary . . . . .	145
<b>7</b>	<b>Online Learning for Interactive Phrase-Based Machine Translation</b>	<b>147</b>
7.1	Introduction . . . . .	147
7.2	Batch Learning versus Online Learning . . . . .	149
7.3	Online Learning as Incremental Learning . . . . .	149
7.3.1	Incremental View of the EM Algorithm . . . . .	150

7.4	Basic IMT System . . . . .	153
7.5	Online IMT System . . . . .	155
7.5.1	Language Model ( $h_1$ ) . . . . .	155
7.5.2	Sentence Length Model ( $h_2$ ) . . . . .	155
7.5.3	Inverse and Direct Phrase-Based Models ( $h_3$ and $h_4$ ) . . . . .	157
7.5.4	Inverse and Direct HMM-Based Alignment Models ( $h_3$ and $h_4$ ) . . . . .	158
7.5.5	Source Phrase Length, Target Phrase Length and Distortion Models ( $h_5$ , $h_6$ and $h_7$ ) . . . . .	162
7.6	Summary . . . . .	162
<b>8</b>	<b>PB-IMT Evaluation</b>	<b>163</b>
8.1	IMT based on Partial Phrase-Based Alignments . . . . .	163
8.1.1	Experiments with the Xerox Corpus . . . . .	164
8.1.2	Experiments with the EU Corpus . . . . .	167
8.2	IMT based on Stochastic Error-Correction Models . . . . .	169
8.2.1	Experiments with the Xerox Corpus . . . . .	169
8.2.2	Experiments with the EU Corpus . . . . .	171
8.3	IMT with Online Learning . . . . .	173
8.3.1	Experiments with the Xerox Corpus . . . . .	173
8.3.2	Experiments with the EU Corpus . . . . .	175
8.4	Summary . . . . .	177
<b>IV</b>	<b>Conclusions and Bibliography</b>	<b>179</b>
<b>9</b>	<b>Conclusions</b>	<b>181</b>
9.1	Summary . . . . .	181
9.2	Scientific Publications . . . . .	185
9.3	Future Work . . . . .	189
	<b>Bibliography</b>	<b>193</b>
<b>V</b>	<b>Appendices</b>	<b>209</b>
<b>A</b>	<b>Incremental EM Algorithm for HMM Alignment Model</b>	<b>211</b>
A.1	Sufficient Statistics . . . . .	211
A.2	E step . . . . .	212
A.3	M step . . . . .	213
<b>B</b>	<b>The Open-Source THOT Toolkit</b>	<b>217</b>
B.1	Design Principles . . . . .	217
B.2	Toolkit Functionality . . . . .	217
B.3	Documentation . . . . .	218
B.4	Public Availability and License . . . . .	218

<b>C</b>	<b>Web-based Interactive Machine Translation Prototype</b>	<b>219</b>
C.1	System Architecture . . . . .	219
C.2	User Interaction Protocol . . . . .	220
C.3	Prototype Functionality . . . . .	220
C.4	Prototype Interface . . . . .	221
<b>D</b>	<b>Symbols and Acronyms</b>	<b>223</b>
D.1	Mathematical Symbols . . . . .	223
D.2	Acronyms . . . . .	225



# LIST OF FIGURES

1.1	Vauquois triangle. . . . .	6
1.2	Architecture of the translation process using the Bayes rule. . . . .	10
1.3	IMT session to translate a Spanish sentence into English. . . . .	25
1.4	An Interactive SMT system. . . . .	25
3.1	Set of consistent bilingual phrases (right) given a word alignment matrix (left). . . . .	41
3.2	Example of the execution of different operations between two alignment matrices. . . . .	41
3.3	Possible bisegmentations for a given word-alignment matrix. . . . .	47
3.4	Tree of possible bisegmentations for a sentence pair. . . . .	49
3.5	Tree of possible bisegmentations including the required information to generate random walks: each edge of the tree is labeled with the number of reachable leafs. . . . .	53
3.6	Example of a file containing sorted counts. . . . .	57
3.7	Bisegmentation example for a sentence pair (left) and the set of values for the hidden variables according to our proposed generative process (right). . . . .	60
4.1	Example of a path in the search graph. The path determines a possible translation of the Spanish source sentence “la casa verde.” along with a valid set of values for the bisegmentation variables $(a_1^K, b_1^K, c_1^K)$ . . . . .	68
4.2	Cache memory architecture. . . . .	87
4.3	An example of the double-trie data structure for the storage of bilingual pairs. The trie at the left stores the source phrases and the one at the right stores the target phrases. . . . .	88
4.4	Example of the expansion of the hypothesis $h$ given $f_1^J \equiv$ “Para ver la lista de recursos” and the target sentence $e_1^I \equiv$ “To view a list of resources”. . . . .	92
5.1	Best- versus breadth-first search comparison executed on the EuTrans-I test corpus. Plots show average translation time per sentence (left) and average score per sentence (right) as a function of the maximum stack size when performing monotonic and non-monotonic translation. . . . .	104
5.2	Best- versus breadth-first search comparison executed on the English-Spanish test set of the Xerox corpus. Plots show average translation time per sentence (left) and average score per sentence (right) as a function of the maximum stack size when performing monotonic and non-monotonic translation. . . . .	105

5.3	Best- versus breadth-first search comparison executed on the Spanish-English test set of the Europarl corpus. Plots show average translation time per sentence (left) and average score per sentence (right) as a function of the maximum stack size when performing monotonic and non-monotonic translation. . . . .	107
5.4	Generalised multiple-stack search experiments executed on the EuTrans-I test corpus. Plots show average translation time (left) and average score (right) per sentence as a function of the granularity (G) parameter. Each curve was obtained using different values of the maximum number of hypotheses stored by the search algorithm ( $L_a$ ). . . . .	108
5.5	Generalised multiple-stack search experiments executed on the English-Spanish test set of the Xerox corpus. Plots show average translation time (left) and average score (right) per sentence as a function of the granularity (G) parameter. Each curve was obtained using different values of the maximum number of hypotheses stored by the search algorithm ( $L_a$ ). . . . .	109
5.6	Generalised multiple-stack search experiments executed on the Spanish-English test set of the Europarl corpus. Plots show average translation time (left) and average score (right) per sentence as a function of the granularity (G) parameter. Each curve was obtained using different values of the maximum number of hypotheses stored by the search algorithm ( $L_a$ ). . . . .	110
5.7	Breadth-first search comparison executed on the EuTrans-I test corpus. Plots show the time cost per sentence in seconds as a function of the average score per sentence using monotonic (left) and non-monotonic (right) translation. Five different functions for mapping hypotheses to stacks were tested. . . . .	111
5.8	Breadth-first search comparison executed on the English-Spanish test set of the Xerox corpus. Plots show the time cost per sentence in seconds as a function of the average score per sentence using monotonic (left) and non-monotonic (right) translation. Five different mapping functions were compared. . . . .	112
5.9	Breadth-first search comparison executed on the Spanish-English test set of the Europarl corpus. Plots show the time cost per sentence in seconds as a function of the average score per sentence using monotonic (left) and non-monotonic (right) translation. Five different mapping functions were tested. . . . .	113
6.1	Example of the expansion of two hypotheses $h_1$ and $h_2$ given $f_1^J \equiv$ “Para ver la lista de recursos” and the user prefix $e_p \equiv$ “To view a”. . . . .	132
6.2	Error-correction model for symbol $a \in \Sigma$ , $\mathcal{A}_a$ . . . . .	137
6.3	Error-correction model for string $\mathbf{x} = x_1x_2x_3$ , $\mathcal{A}_x$ . The model has been obtained by concatenating $\mathcal{A}_{x_1}$ , $\mathcal{A}_{x_2}$ and $\mathcal{A}_{x_3}$ . . . . .	137
6.4	Reduced version of $\mathcal{A}_x$ . . . . .	137
6.5	Example of how the IMT suffix is determined in our alternative IMT formalisation. . . . .	140
6.6	Error-correction model based on PFSMs for IMT given the sentence $e_1^I$ : $\mathcal{B}_{e_1^I}$ . The states of the PFSM are labelled with the words of the target sentence $e_1^I$ . . . . .	141
7.1	An Online Interactive SMT system. . . . .	148



8.1	KSMR evolution translating a portion of the Xerox training corpora. A monotonic online IMT system with log-linear weights tuned via MERT was used. .	174
C.1	IMT system architecture. . . . .	220
C.2	Translating documents with the proposed system. . . . .	221
C.3	Index of available corpora. . . . .	222
C.4	Prototype interface. The source text segments are automatically extracted from source document. . . . .	222



# LIST OF TABLES

1.1	EuTrans-I corpus statistics. . . . .	29
1.2	Europarl corpus statistics for three different language pairs. . . . .	29
1.3	Europarl language model data. . . . .	30
1.4	Hansards corpus statistics. . . . .	30
1.5	Xerox corpus statistics for three different language pairs. . . . .	31
1.6	EU corpus statistics for three different language pairs. . . . .	31
3.1	Set of all monolingual segmentations for a source sentence of 4 words and their representation as a binary number of 3 bits. . . . .	44
3.2	Bilingual phrase counts and fractional counts for RF and BR estimation, respectively, for the sentence pair shown in Figure 3.1. . . . .	47
3.3	Statistics of different phrase models estimated from the Europarl corpus ranging over the maximum phrase size (denoted by $\mathbf{m}$ ). . . . .	56
4.1	Values returned by the $\mu_1$ and $\mu_2$ function defined as a composition of the $\alpha$ and $\beta$ functions. . . . .	81
5.1	Statistics of both conventional estimation and fragment-by-fragment estimation for the English-Spanish Europarl corpus and different values of the maximum phrase size. The statistics include the time in seconds and the main memory size in GBytes required by the estimation process. . . . .	102
5.2	Description of the log-linear model used in the experiments. . . . .	103
5.3	Influence of different models on the translation quality measured in terms of BLEU. The results were obtained using the Spanish-English Europarl test corpus (In+Out) and its in-domain (In) and out-domain (Out) subsets. A breadth-first multiple-stack algorithm was used to generate the translations. MERT was used to adjust the weights of the log-linear model. . . . .	114
5.4	BLEU results obtained with the Europarl test corpora when translating from Spanish, French and German to the English language. The translations were generated by means of a breadth-first multiple-stack algorithm. Monotonic and non-monotonic search were used. The weights of the log-linear combination were tuned via MERT. . . . .	115
5.5	Comparison of translation quality (measured according to BLEU) between Moses and our proposed translation system (BB-ALG) when translating the Europarl test corpora. 95% confidence intervals are shown. . . . .	115

5.6	Comparison of translation quality (measured according to BLEU) between Moses and our proposed translation system (BB-ALG) when translating the Europarl test corpora. The log-linear model used by Moses did not include lexical components. 95% confidence intervals are shown. . . . .	116
5.7	BLEU results when translating the Xerox test corpora from English to Spanish, French and German. Translations were generated by means of a breadth-first multiple-stack algorithm. MERT was used to tune the weights of the log-linear model. . . . .	116
5.8	Comparison of translation quality (measured according to BLEU) between Moses and our proposed translation system (BB-ALG) when translating the Xerox test corpora. 95% confidence intervals are shown. . . . .	117
5.9	Translation quality results measured in terms of BLEU when translating the EU test corpora from Spanish, French and German to the English language. A breadth-first multiple-stack algorithm was used to generate the translations. The weights of the log-linear combination were adjusted via the MERT algorithm. . . . .	117
5.10	Comparison of translation quality (measured according to BLEU) between Moses and our proposed translation system (BB-ALG) when translating the EU test corpora. 95% confidence intervals are shown. . . . .	117
5.11	Comparison of estimation time cost in seconds using RF and BR estimation when translating from English to Spanish with the EuTrans-I, Xerox and Europarl corpora. . . . .	118
5.12	Comparison of translation quality (measured according to BLEU) using RF and BR estimation when translating from English to Spanish with the EuTrans-I, Xerox and Europarl corpora. 95% confidence intervals are shown. . . . .	119
5.13	Number of phrases, disk accesses, total time (in secs), and disk overhead required to retrieve the translations for the phrases of the Spanish-English Europarl test set, ranging over the value of $\alpha$ . . . . .	120
5.14	Number of queries, % of cache misses, total, per sentence and per query locating time (in secs.) required by all model queries when translating the Spanish-English Europarl test set ( $\alpha = 100$ constitutes the baseline system). A breadth-first multiple-stack algorithm was used to generate the translations. Such algorithm implemented monotonic search. Default log-linear weights were used. . . . .	121
5.15	Comparative alignment quality results (in %) using different smoothing techniques for NO-NUL and NUL alignments. A breadth-first multiple-stack algorithm was used to generate the alignments. Such algorithm implemented non-monotonic search. Default log-linear weights were used. Best results are shown in bold. . . . .	123
5.16	Alignment quality results (in %) using GT+LEX <sub>BO</sub> smoothing for NO-NUL and NUL alignments. A breadth-first multiple-stack algorithm was used to generate the alignments. The search algorithm implemented non-monotonic search. Log-linear weights were tuned using MERT. . . . .	123
6.1	Summary of applications of the generalised formalisation. . . . .	144

8.1	KSMR results for the three Xerox corpora (for both direct and inverse translation directions separated by the symbol “/”) for a monotonic IMT system and different smoothing techniques. Geometric distributions were selected to implement the $h_5$ and $h_6$ feature functions. Default weights for the log-linear model were used. Best results are shown in bold. . . . .	164
8.2	KSMR results for the three Xerox corpora (for both direct and inverse translation directions separated by the symbol “/”) for all possible combinations of the probability distributions for the $h_5$ and $h_6$ feature functions when using two different smoothing techniques. A monotonic IMT system with default log-linear model weights were used. Best results are shown in bold. . . . .	165
8.3	KSMR results for the three Xerox corpora, using a monotonic IMT system with three different smoothing techniques. Geometric distributions were used to implement the $h_5$ and $h_6$ feature functions. MERT was performed. The average time in seconds per interaction is also reported. . . . .	166
8.4	KSMR results for the three Xerox corpora, using a non-monotonic IMT system with GT+LEX <sub>BO</sub> smoothing. Geometric distributions were used to implement the $h_5$ and $h_6$ feature functions. MERT was performed. The average time in seconds per interaction is also reported. . . . .	166
8.5	CER and PKSR obtained with the post-editing approach and KSR obtained with our proposed IMT system based on smoothing techniques (a monotonic IMT system with GT+LEX <sub>BO</sub> smoothing tuned with MERT was used, geometric distributions were selected to implement the $h_5$ and $h_6$ feature functions). The results were obtained for the Xerox corpora. . . . .	167
8.6	KSMR results comparison of our IMT system based on partial statistical phrase-based alignments (a monotonic IMT system with GT+LEX <sub>BO</sub> smoothing tuned with MERT was used, geometric distributions were selected to implement the $h_5$ and $h_6$ feature functions) and three different state-of-the art IMT systems. 95% confidence intervals are shown. The experiments were executed on the Xerox corpora. Best results are shown in bold. . . . .	167
8.7	KSMR results for the three EU corpora, using a monotonic IMT system with GT+LEX <sub>BO</sub> smoothing. Geometric distributions were used to implement the $h_5$ and $h_6$ feature functions. MERT was performed. The average time in seconds per interaction is also reported. . . . .	168
8.8	CER and PKSR obtained with the post-editing approach and KSR obtained with our proposed IMT system based on smoothing techniques (a monotonic IMT system with GT+LEX <sub>BO</sub> smoothing tuned with MERT was used, geometric distributions were selected to implement the $h_5$ and $h_6$ feature functions). The results were obtained for the EU corpora. . . . .	168
8.9	KSMR results comparison of our IMT system based on partial statistical phrase-based alignments (a monotonic IMT system with GT+LEX <sub>BO</sub> smoothing tuned with MERT was used, geometric distributions were selected to implement the $h_5$ and $h_6$ feature functions) and three different state-of-the art IMT systems. 95% confidence intervals are shown. The experiments were executed on the EU corpora. Best results are shown in bold. . . . .	169

8.10	KSMR results for the three Xerox corpora, using an IMT system based on stochastic error-correction models. Word graphs were generated by means of a monotonic SMT system. MERT was performed. The average time in seconds per interaction is also reported. . . . .	170
8.11	KSMR results for the three Xerox corpora, using an IMT system based on stochastic error-correction models. Word graphs were generated by means of a non-monotonic SMT system. MERT was performed. The average time in seconds per interaction is also reported. . . . .	170
8.12	CER and PKSR obtained with the post-editing approach and KSR obtained with our proposed IMT system based on error-correction models (word graphs were generated by means of a monotonic SMT system). The results were obtained for the Xerox corpora. . . . .	171
8.13	KSMR results comparison of our IMT system based on stochastic error-correction models and four different state-of-the art IMT systems (word graphs were generated by means of a monotonic SMT system). 95% confidence intervals are shown. The experiments were executed on the Xerox corpora. Best results are shown in bold. . . . .	171
8.14	KSMR results for the three EU corpora, using an IMT system based on stochastic error-correction models. Word graphs were generated by means of a monotonic SMT system. MERT was performed. The average time in seconds per iteration is also reported. . . . .	172
8.15	CER and PKSR obtained with the post-editing approach and KSR obtained with our proposed IMT system based on error-correction models (word graphs were generated by means of a monotonic SMT system). The results were obtained for the EU corpora. . . . .	172
8.16	KSMR results comparison of our IMT system based on stochastic error-correction models and four different state-of-the art IMT systems (word graphs were generated by means of a monotonic SMT system). 95% confidence intervals are shown. The experiments were executed on the EU corpora. Best results are shown in bold. . . . .	172
8.17	BLEU and KSMR results for the Xerox test corpora using the batch and the online IMT systems. Both systems used monotonic search with log-linear weights tuned via MERT. The average online learning time (LT) in seconds is shown for the online system. . . . .	175
8.18	KSMR results comparison of our system and three different state-of-the-art batch systems. The experiments were executed on the Xerox corpora. Best results are shown in bold. . . . .	176
8.19	BLEU and KSMR results for the French-English EU test corpus using the batch and the online IMT systems. Both IMT systems used monotonic search. MERT was performed. The average online learning time (LT) in seconds is shown for the online system. . . . .	176

8.20 BLEU and KSMR results for an alternative partition of the French-English EU corpus using the batch and the online IMT systems. Both systems used monotonic search. The log-linear weights were tuned by means of the MERT algorithm. The average online learning time (LT) in seconds is shown for the online system. . . . . 177





# LIST OF ALGORITHMS

3.1	Pseudocode for the <code>phrase_extract</code> algorithm. . . . .	43
3.2	Pseudocode for the <code>brf_phr_extract</code> algorithm. . . . .	48
3.3	Pseudocode for the <code>brf_phr_extract_dp</code> algorithm. . . . .	50
3.4	Pseudocode for the <code>bisegm_random_walk</code> algorithm. . . . .	54
3.5	Pseudocode for the <code>frag_by_frag_training</code> algorithm. . . . .	57
3.6	Pseudocode for the <code>merge_counts</code> algorithm. . . . .	58
4.1	Pseudocode for the <code>dp_search</code> algorithm. . . . .	71
4.2	Pseudocode for the <code>bb_search</code> algorithm. . . . .	74
4.3	Pseudocode for the <code>expand</code> algorithm. . . . .	75
4.4	Pseudocode for the <code>push_rec</code> algorithm. . . . .	76
4.5	Pseudocode for the <code>mon_expand</code> algorithm. . . . .	77
4.6	Pseudocode for the <code>phralig_expand</code> algorithm. . . . .	93
6.1	Pseudocode for the <code>imt_expand</code> algorithm. . . . .	134
7.1	Pseudocode for the <code>update_suff_stats_lm</code> algorithm. . . . .	156
7.2	Pseudocode for the <code>updD</code> algorithm. . . . .	156