

# List of Tables

3.1	Value of $g(w_{i-1}, w_i)$ and $g(w_{i-2}, w_{i-1}, w_i)$ . . . . .	40
3.2	Value of $g(w_{i-1}, w_i)$ and $g(t, w_i)$ . . . . .	54
3.3	All $2^M$ intersections of $M$ subsets. . . . .	57
3.4	The mapping between intersections and binary numbers. $I_k$ is the intersection of subsets with value 1 under them. $k$ is a binary with value 1 in corresponding bits. . . . .	58
3.5	Comparison of Switchboard and Broadcast News. . . . .	79
3.6	Number of operations and nominal speed-up of trigram models . . . .	81
3.7	Number of operations and nominal speed-up of four-gram models . .	81
3.8	Running time (in CPU-Hours) for ME trigram models . . . . .	82
3.9	Number of operations and nominal speed-up for syntactic models . .	82
3.10	Training time (in CPU-hours) for syntactic models . . . . .	83
3.11	Number of operations and nominal speed-up for topic models . . . . .	83
3.12	Running time (in CPU-Hours) for topic models . . . . .	84
3.13	Training time for composite model in Switchboard . . . . .	85
4.1	Sensitive words for the topic “CLOTHES”. . . . .	89
4.2	Perplexity and WER of back-off trigram models and an ME model with the same constraints . . . . .	95
4.3	Topic assignment based on erroneous recognizer hypotheses causes little degradation in performance. . . . .	96
4.4	Topic assignments of test conversations using different methods. . . .	97
4.5	Topic assignments of test conversations using different methods (Cont). .	98
4.6	Dynamic topic assignment for individual utterances based on the current and four preceding utterances. . . . .	101
4.7	Topic dynamics viewed through (dis)agreement of utterance-level and conversation-level topic assignment. . . . .	101
4.8	Analysis of performance gains from the topic-dependent model. . . .	102
4.9	Performance improvement on content-bearing words. . . . .	102
4.10	Comparison with 70 interpolated topic N-gram models. . . . .	104

4.11	Perplexity and WER of a cache-based model and a maximum entropy model with topic constraints. . . . .	105
4.12	Repeated errors vs. non-repeated errors. . . . .	105
4.13	Perplexity and WER of language models for the BN corpus . . . . .	107
4.14	Comparison with 100 interpolated topic N-gram models. . . . .	108
5.1	The effect of dropping head-word constraints when they coincide with the two preceding words. . . . .	118
5.2	Dependence on non-terminals (NT). . . . .	119
5.3	Dependence on components of the syntactic heads: non-terminals (NT), head-words (HW) and both. . . . .	119
5.4	Splitting results: $h_{i-2}, h_{i-1} = w_{i-2}, w_{i-1}$ vs $h_{i-2}, h_{i-1} \neq w_{i-2}, w_{i-1}$ . . .	121
5.5	Conditioning on non-terminal (NT) vs part-of-speech (POS) tags: effect of syntactic analysis. . . . .	123
5.6	Trigram coverage: $h_{i-2}, h_{i-1} = w_{i-2}, w_{i-1}$ vs $h_{i-2}, h_{i-1} \neq w_{i-2}, w_{i-1}$ . .	123
5.7	ME v/s interpolated syntactic models (cf. Chelba & Jelinek (1999)).	124
5.8	Baseline perplexities and WERs for BN. . . . .	126
5.9	Performance of syntactic models for BN. . . . .	126
5.10	Memory requirement and training time per iteration for SWBD and BN. The data in the last row are estimated numbers. . . . .	128
6.1	Performance of ME language models with N-gram, head-word (HW), non-terminal label (NT) and topic dependencies. . . . .	133
6.2	Perplexity and WER for content words and stop words. . . . .	135
6.3	Percentage of content/stopwords with syntactic constraints. . . . .	135
6.4	Performance improvement on different history classes. . . . .	136
6.5	Proportion of test data of different word classes and/or history classes.	137
6.6	WER based on different history classes and word classes. . . . .	137
6.7	Coverage of trigrams of different word classes and/or history classes. .	138
6.8	Performance of ME language models with N-gram, syntactic and topic dependencies. . . . .	139
7.1	ARPA format for back-off models. . . . .	143
7.2	Speed-up of the pre-nominalized model. . . . .	147
7.3	Perplexity and WER of using the BO and ME models in the first pass recognition. . . . .	148
7.4	Influence of approximation on topic models. . . . .	150
7.5	Hit rate for different models. . . . .	156
8.1	Nominal speed-up in training Charniak parser. . . . .	161
8.2	Summary of experimental results (SWBD and 14M BN). . . . .	162

A.1	Significance of WER difference between language models (SWBD). Negative values in Diff. WER indicate that Model B is better than Model A. . . . .	167
A.2	Significance of WER difference between language models (BN). Nega- tive values in Diff. WER indicate that Model B is better than Model A. . . . .	168
C.1	Memory requirement for ME models. . . . .	182
C.2	Content of Directories. . . . .	184