

Contents

1	Introduction	15
1.1	Preliminary Remarks	15
1.2	Central Claims	17
1.3	Motivation	18
1.4	Overview of the Thesis	26
1.5	Published Work	28
2	Probabilistic Context-Free Grammars for Syllabification and Grapheme-to-Phoneme Conversion	29
2.1	Introduction	29
2.2	Method	30
2.3	Experiment 1: Syllabification	30
2.4	Experiment 2: Grapheme-to-Phoneme Conversion	36
2.5	Discussion	42
2.6	Conclusion	44
3	Inducing Probabilistic Syllable Classes Using Multivariate Clustering	45
3.1	Introduction	45
3.2	Method	47
3.3	Experiment 3: Three-dimensional Clustering	50
3.4	Experiment 4: Five-dimensional Clustering	57
3.5	Experiment 5 and 6: Three- and Five-dimensional Type-Based Models	64
3.6	Discussion	67
3.7	Conclusion	69

4 Automatic Detection of Syllable Boundaries Combining Treebank and Bracketed Corpora Training	71
4.1 Introduction	71
4.2 Method	72
4.3 Experiment 7: Word Token Training	77
4.4 Experiment 8: Word Type Training	88
4.5 Discussion	90
4.6 Conclusion	92
5 Probabilistic Context-Free Grammars for Phonology	95
5.1 Introduction	95
5.2 Method	96
5.3 Experiment 9: Multilingual Modeling of Syllable Structure	98
5.4 Evaluation	99
5.5 Discussion	104
5.6 Conclusion	107
6 Conclusion	125
A German 5-dimensional Clustering Model	137
B English 5-dimensional Clustering Model	155
C Advanced Positional Syllable Structure Grammar	175