

# **Part-of-speech Bootstrapping using Lexically-Specific Frames**

Richard Eduard Leibbrandt  
B.Sc. (Hons), B.Soc.Sci. (Hons), M.Sc.

School of Computer Science, Engineering and Mathematics  
Faculty of Science and Engineering  
Flinders University of South Australia  
July 2009

Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

<b>1 INTRODUCTION .....</b>	<b>10</b>
THE PROBLEM OF PART-OF-SPEECH INDUCTION .....	10
1.1 OUTLINE OF THIS THESIS .....	11
<b>2 PARTS-OF-SPEECH AND THEIR ROLE IN GRAMMAR.....</b>	<b>13</b>
2.1 INTRODUCTION.....	13
2.2 PARTS-OF-SPEECH IN TRADITIONAL LINGUISTIC THEORY .....	14
2.2.1 <i>Open word classes</i> .....	15
2.2.2 <i>Closed word classes</i> .....	18
2.3 THE ROLE OF CONSTRUCTIONS IN ASSIGNING PARTS-OF-SPEECH .....	20
2.3.1 <i>Transformational Generative Grammar</i> .....	20
2.3.2 <i>Cognitive Grammar and Construction Grammar</i> .....	21
2.4 A CLOSER LOOK AT OPEN WORD CLASSES .....	23
2.4.1 <i>Langacker's Cognitive Grammar</i> .....	24
2.4.2 <i>Croft's Radical Construction Grammar</i> .....	25
2.4.3 <i>The universal status of the main open classes</i> .....	27
2.4.4 <i>Constructions as clues to word meaning</i> .....	28
<b>3 CHILDREN'S KNOWLEDGE AND DISCOVERY OF PARTS-OF-SPEECH .....</b>	<b>31</b>
3.1 INTRODUCTION.....	31
3.2 BOOTSTRAPPING THEORIES .....	31
3.2.1 <i>Semantic bootstrapping</i> .....	32
3.2.2 <i>Distributional bootstrapping</i> .....	33
3.2.3 <i>Other bootstrapping theories</i> .....	36
3.3 CHILDREN'S EARLY USE OF LEXICALLY-SPECIFIC FRAMES .....	37
3.3.1 <i>Braine's pivot grammars</i> .....	38
3.3.2 <i>Tomasello's Verb Island hypothesis</i> .....	38
3.3.3 <i>Lieven, Pine and colleagues: lexically-specific frames</i> .....	39
3.4 CHILDREN'S UNDERSTANDING OF CONSTRUCTIONS AND FRAMES AS CUES TO WORD CLASSES.....	41
3.4.1 <i>Nouns</i> .....	43
3.4.2 <i>Verbs</i> .....	48
3.4.3 <i>Adjectives</i> .....	51
3.5 THE INTERACTION BETWEEN CONSTRUCTIONS AND EMBEDDED WORDS.....	54
3.6 THE ROLE OF FUNCTION WORDS IN LEXICALLY-SPECIFIC FRAMES .....	57
3.6.1 <i>Early knowledge of function words</i> .....	57
3.6.2 <i>Early knowledge of the relationship between function words and other words</i> .....	60
3.6.3 <i>How do lexically-specific frames develop?</i> .....	64
3.6.4 <i>Artificial and other language learning experiments</i> .....	68
3.7 SYNTHESIS: A POSSIBLE ROLE FOR LEXICALLY-SPECIFIC FRAMES IN PART-OF-SPEECH BOOTSTRAPPING ...	72
<b>4 COMPUTATIONAL MODELS OF AUTOMATIC PART-OF-SPEECH INDUCTION .....</b>	<b>74</b>
4.1 LEXICAL CATEGORIZATION FROM CONTEXTUAL DISTRIBUTION .....	74
4.1.1 <i>Finch</i> .....	74
4.1.2 <i>Redington, Chater and Finch</i> .....	76
4.1.3 <i>Mintz, Newport and Bever</i> .....	77
4.1.4 <i>Clark</i> .....	78
4.1.5 <i>Frequent Frames</i> .....	79
4.1.6 <i>Schütze</i> .....	81
4.1.7 <i>Cartwright &amp; Brent</i> .....	81
4.2 SYNTAX LEARNING.....	82

4.2.1	<i>EMILE</i>	82
4.2.2	<i>ABL</i>	85
4.2.3	<i>SNPR/ICMAUS</i>	88
4.2.4	<i>GRIDS</i>	90
4.2.5	<i>ADIOS</i>	91
4.2.6	<i>MOSAIC</i>	92
4.2.7	<i>Powers</i>	93
4.3	OTHER MODELS	95
4.3.1	<i>Yuret</i>	95
4.3.2	<i>PARSER</i>	97
4.4	COMPARISON OF COMPUTATIONAL MODELS	97
4.4.1	<i>Syntagmatic patterns</i>	98
4.4.2	<i>Paradigmatic categories</i>	99
4.4.3	<i>The three main parts-of-speech</i>	99
4.4.4	<i>Word type ambiguity</i>	100
4.4.5	<i>Contextual ambiguity</i>	101
4.4.6	<i>Contexts as explicit linguistic units</i>	102
4.4.7	<i>Lexically-specific contexts</i>	102
4.4.8	<i>The current work</i>	103
<b>5 THE APPROACH BASED ON LEXICALLY-SPECIFIC FRAMES</b>		<b>106</b>
5.1	INTRODUCTION	106
5.2	RATIONALE FOR THE FRAME APPROACH	107
5.2.1	<i>The formation of construction frames</i>	107
5.2.2	<i>Amalgamation of frames into categories</i>	108
5.3	BASIC APPROACH	112
5.3.1	<i>Syntagmatic contextual patterns</i>	115
5.3.2	<i>Paradigmatic parts-of-speech</i>	116
5.4	METHODOLOGICAL ISSUES AND ASSUMPTIONS	117
5.4.1	<i>Corpus and preprocessing</i>	117
5.4.2	<i>Segmentation into orthographic words</i>	118
5.5	EVALUATION OF CLUSTERING RESULTS	119
5.5.1	<i>Quantitative evaluation against a gold standard</i>	119
5.5.2	<i>Quantitative evaluation measures</i>	120
<b>6 FULL-UTTERANCE FRAMES</b>		<b>132</b>
6.1	INTRODUCTION	132
6.2	AUTOMATIC DISCOVERY PROCEDURE	132
6.2.1	<i>Outline</i>	132
6.2.2	<i>Details</i>	134
6.3	PSYCHOLOGICAL CONSIDERATIONS	136
6.4	PRELIMINARY CONSIDERATIONS	139
6.4.1	<i>Variant treatments of the frequent/infrequent word dichotomy</i>	139
6.4.2	<i>Clustering analysis</i>	140
6.4.3	<i>The number of candidate frame-building words</i>	144
6.5	IMPLEMENTATION	145
6.5.1	<i>Qualitative results</i>	145
6.5.2	<i>Clustering results</i>	151
6.5.3	<i>Effect of number of frequent words used</i>	158
6.5.4	<i>Quantitative evaluation</i>	162
6.6	DISCUSSION	166

<b>7</b>	<b>RESOLVING AMBIGUITY WITH CO-CLUSTERING.....</b>	<b>168</b>
7.1	INTRODUCTION.....	168
7.1.1	<i>Constraints on the representation of frame and word ambiguity</i> .....	168
7.1.2	<i>Co-clustering and biclustering</i> .....	170
7.2	FUZZY CO-CLUSTERING.....	172
7.2.1	<i>Procedure</i> .....	172
7.2.2	<i>Psychological considerations</i> .....	176
7.3	DISCRETE CO-CLUSTERING ALGORITHMS.....	179
7.3.1	<i>Seed words and frames</i> .....	180
7.3.2	<i>Allocation matrices</i> .....	183
7.3.3	<i>Parsimony-driven co-clustering</i> .....	183
7.3.4	<i>Conflict-driven co-clustering</i> .....	190
7.4	EVALUATION.....	196
7.5	OTHER ISSUES.....	203
7.5.1	<i>The robustness of the three main categories in the co-clustering approaches</i> .....	203
7.5.2	<i>Evaluation of independent frame and word ambiguity</i> .....	210
7.5.3	<i>Frames with multiple-X sequences</i> .....	213
7.6	DISCUSSION.....	218
<b>8</b>	<b>HIERARCHICALLY NESTED FRAMES.....</b>	<b>220</b>
8.1	INTRODUCTION.....	220
8.2	FRAME DISCOVERY PROCEDURE.....	221
8.3	PSYCHOLOGICAL CONSIDERATIONS.....	225
8.4	IMPLEMENTATION.....	228
8.5	QUALITATIVE RESULTS.....	234
8.6	QUANTITATIVE RESULTS.....	238
8.7	A COMPARISON WITH CAMERON-FAULKNER ET AL. (2003).....	246
8.7.1	<i>Differences from the current work</i> .....	246
8.7.2	<i>Comparison of frames</i> .....	249
8.8	DISCUSSION.....	259
<b>9</b>	<b>PREDICTION-BASED FRAMES.....</b>	<b>261</b>
9.1	INTRODUCTION.....	261
9.2	MISCONCEPTIONS ABOUT PREDICTABILITY.....	262
9.3	PREDICTIVE RELATIONS AS INDICATORS OF FRAME-FILLER RELATIONS.....	266
9.4	ASSEMBLING COMPLEX FRAMES FROM INDIVIDUAL ASSOCIATIVE LINKS.....	269
9.4.1	<i>Basic frame patterns</i> .....	270
9.5	PSYCHOLOGICAL CONSIDERATIONS.....	274
9.6	FRAME DISCOVERY PROCEDURE.....	276
9.6.1	<i>Obtaining frame probabilities</i> .....	276
9.6.2	<i>Parsing the corpus for frames</i> .....	278
9.7	IMPLEMENTATION.....	280
9.7.1	<i>Qualitative results</i> .....	283
9.7.2	<i>Quantitative results</i> .....	286
9.8	A COMPARISON WITH CAMERON-FAULKNER ET AL. (2003).....	288
9.9	PROBLEMS WITH PREDICTION-BASED FRAMES.....	289
9.10	DISCUSSION.....	290
<b>10</b>	<b>SUMMARY AND COMPARISON TO OTHER COMPUTATIONAL MODELS.....</b>	<b>291</b>
10.1	SUMMARY OF EMPIRICAL RESULTS.....	291
10.2	DIFFERENCES IN THE METHOD OF CONTEXT DISCOVERY.....	296

10.2.1	<i>The nested frame approach compared to ABL</i> .....	296
10.2.2	<i>The full-utterance frame approach compared to EMILE</i> .....	300
10.2.3	<i>The current approach compared to Frequent Frames</i> .....	301
10.2.4	<i>The frame approach compared to Yuret (1998)</i> .....	304
10.3	DIFFERENCES IN THE METHOD OF FOCAL WORD CATEGORIZATION.....	305
10.3.1	<i>Clustering in EMILE</i> .....	305
10.3.2	<i>Clustering in Frequent Frames</i> .....	307
10.3.3	<i>Criticisms of category formation via clustering</i> .....	308
<b>11</b>	<b>FUTURE EXTENSIONS OF THE PART-OF-SPEECH DISCOVERY FRAMEWORK</b> .....	<b>313</b>
11.1	EXTENDING DISTRIBUTIONAL BOOTSTRAPPING.....	313
11.1.1	<i>Extended coverage of the corpus: generalizing beyond the current data set</i> .....	313
11.1.2	<i>Heuristics about the structure of frames</i> .....	313
11.1.3	<i>Morphological awareness</i> .....	315
11.1.4	<i>Finding functional categories</i> .....	315
11.1.5	<i>Extending the prediction-based approach</i> .....	315
11.1.6	<i>Other languages</i> .....	316
11.2	SEMANTIC ASPECTS: MICROJAEA.....	316
11.2.1	<i>The MicroJaea system</i> .....	318
11.2.2	<i>Scripting language</i> .....	318
11.2.3	<i>A framework for language learning experiments</i> .....	320
	<b>APPENDIX 1: PREPROCESSING PERFORMED ON THE MANCHESTER CORPUS</b> .....	<b>330</b>
	<b>APPENDIX 2: FULL COMPARISON OF CURRENT FULL-UTTERANCE AND NESTED FRAMES AGAINST THE FRAMES IDENTIFIED BY CAMERON-FAULKNER ET AL. (2003)</b> .....	<b>335</b>
	<b>REFERENCES</b> .....	<b>387</b>

## Abstract

The work in this thesis presents and evaluates a number of strategies by which English-learning children might discover the major open-class parts-of-speech in English (nouns, verbs and adjectives) on the basis of purely distributional information. Previous work has shown that parts-of-speech can be readily induced from the distributional patterns in which words occur. The research reported in this thesis extends and improves on this previous work in two major ways, related to the *constructional* status of the utterance contexts used for distributional analysis, and to the way in which previous studies have dealt with categorial *ambiguity*.

Previous studies that have induced parts-of-speech from word distributions have done so on the basis of fixed “windows” of words that occur before and after the word in focus. These contexts are often not constructions of the language in question, and hence have dubious status as elements of linguistic knowledge. A great deal of recent evidence (e.g. Lieven, Pine & Baldwin, 1997; Tomasello, 1992) has suggested that children’s early language may be organized around a number of lexically-specific constructional frames with slots, such as “a X”, “you X it”, “draw X on X”. The work presented here investigates the possibility that constructions such as these may be a more appropriate domain for the distributional induction of parts-of-speech. This would open up the possibility of a treatment of part-of-speech induction that is more closely integrated with the acquisition of syntax.

Three strategies to discover lexically-specific frames in the speech input to children are presented. Two of these strategies are based on the interplay between more and less frequent words in English utterances: the more frequent words, which are typically function words or light verbs, are taken to provide the schematic “backbone” of an utterance. In the first strategy, all frames are schematic structures for full utterances. The second strategy extends this approach to include multi-word sequences that frequently occur embedded inside other frequent multi-word sequences. The third strategy is based around pairs of words in which the occurrence of one word is highly predictable from that of the other, but not vice versa; from these basic slot-filler relationships, larger frames are assembled.

These techniques were implemented computationally and applied to a corpus of child-directed speech. Each technique yielded a large set of lexically-specific frames, many of which could plausibly be regarded as constructions. In a comparison with a manual analysis of the same corpus by Cameron-Faulkner, Lieven and Tomasello (2003), it is shown that most of the constructional frames identified in the manual analysis were also produced by the automatic techniques.

After the identification of potential constructional frames, parts-of-speech were formed from the patterns of co-occurrence of words in particular constructions, by means of hierarchical clustering. The resulting clusters produced are shown to be quite similar to the major English parts-of-speech of nouns, verbs and adjectives. Each individual word token was assigned a part-of-speech on the basis of its constructional context. This

categorization was evaluated empirically against the part-of-speech assigned to the word in question in the original corpus. The resulting categorization is shown to be, to a great extent, in agreement with the manual categorization.

These strategies deal with the categorial ambiguity of words, by allowing the frame context to determine part-of-speech. However, many of the frames produced were themselves ambiguous cues to part-of-speech. For this reason, strategies are presented to deal with both word and context ambiguity. Three such strategies are proposed. One considers membership of a part-of-speech to be a matter of *degree* for both word and contextual frame. A second strategy attempts to discretely assign multiple parts-of-speech to words and constructions in a way that imposes internal *consistency* in the corpus. The third strategy attempts to assign only the minimally-required multiple categories to words and constructions so as to provide a *parsimonious* description of the data.

Each of these techniques was implemented and applied to each of the three frame discovery techniques, thereby providing category information about *both* the frame and the word. The subsequent assignment of parts-of-speech was done by combining word and frame information, and is shown to be far more accurate than the categorization based on frames alone. This approach can be regarded as addressing certain objections against distributional part-of-speech bootstrapping that have been raised by Pinker (1979, 1984, 1987).

Lastly, a framework for extending this research is outlined that allows semantic information to be incorporated into the process of category induction.

I certify that this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

---

Richard Eduard Leibbrandt

21 July 2009



## Acknowledgements

A number of people have helped to make the eventual completion of this thesis possible.

I would like to express my thanks first of all to my thesis supervisor David Powers, for his friendship and encouragement throughout these last few years. David gave me free rein to develop my own ideas, engaged with me on the scientific issues while respecting my viewpoint when we disagreed, served as an inexhaustible source of technical (and sometimes esoteric) information, and helped me to develop into a scientific researcher.

To my friends and colleagues in the Artificial Intelligence Laboratory and the School of Computer Science, Engineering and Mathematics at Flinders University, Graham Bignell, Aaron Ceglar, Denise de Vries, Sean Fitzgibbon, Shawn Hagggett, Trent Lewis, Martin Luerksen, Takeshi Matsumoto, Anna Shillabeer, Kenneth Treharne and Dongqiang Yang, I give my thanks for helping to create a welcoming and supportive community in which to work.

Thanks to Darius Pfitzner for your friendship through almost five years of despair, gnawing self-doubt, etc., etc. while we struggled with our respective PhDs, having interminable cuppas and spending many afternoons and evenings enjoying meals at your home (with the additional forbearance of Susan, Mellion and Pheobe). Thank you for showing me the value of sticking at it and pushing through, and that you can achieve more if you attempt more. We dunnit mate.

Lastly, thank you to my parents for supporting me financially during my undergraduate studies. This is the culmination of that process (I promise. It's the last one you can get).