

NECESSARY BIAS IN NATURAL LANGUAGE LEARNING

By

Lisa Sue Pearl

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:
Professor Amy Weinberg, Chair
Professor Jeffrey Lidz
Professor William Idsardi
Professor Charles Yang
Professor James Reggia

© Copyright by
Lisa Sue Pearl
2007

Dedications

To Amy Weinberg, who put up with a ridiculous number of last minute questions and drafts of *everything*. With good humor, to boot.

To Norbert Hornstein, who was always exceptional at convincing me that my ideas were any good. He's really quite persuasive.

To Charles Yang, whose work inspired all of this. More than once.

To Jeff Lidz, who helped me figure out how to write things that were actually comprehensible and who was always full of even more ideas.

To Bill Idsardi, who was also always full of even more ideas and who made me think very carefully about all my mathematical work.

To Philip Resnik, whose boundless energy and enthusiasm and questions completely inspired me time and time again.

To David Lightfoot, who taught the first theoretical linguistics class I ever took and who I wanted to grow up to be.

To Michelle Hugue, who taught me how to make computation into a tool and about how there are so many different ways to measure success.

To Peggy Antonisse, whose thoroughly sensible advice and calming influence were welcome balm in times of great stress and panic.

Acknowledgements

This work would have been a long time in coming without the listening ears, unwavering support, and good advice of many, many people, among them: Amy Weinberg, Norbert Hornstein, Charles Yang, Jeff Lidz, Bill Idsardi, Philip Resnik, David Lightfoot, Michelle Hugue, Peggy Antonisse, Colin Phillips, Andrea Zukowski, Heather Taylor, Brian Dillon, and Ivano Caponigro. This work was additionally supported by a National Science Foundation Graduate Research Fellowship.

Table of Contents

Dedications	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	viii
List of Figures	ix
Chapter 1: A Theory of the Language Learning Mechanism	1
1.1 The Mechanism of Language Learning	1
1.2 Language Development: Constraints on the Hypothesis Space	1
1.3 Formalizing the Language Acquisition Mechanism	2
1.4 Domain Specificity and Domain Generality	2
1.5 Investigating the Components of the Learning Framework	3
1.6 Computational Investigations of Data Intake Filtering	4
1.7 Organization of Dissertation	4
Chapter 2: Bayesian Updating in a Linguistic Framework	6
2.1 Bayesian Updating: Overview	6
2.1.1 A Simple Case: Two Non-overlapping Hypotheses, Equally Likely	6
2.1.2. A Variant on the Simple Case: Two Non-overlapping Hypotheses, with an Initial Bias for One Hypothesis	8
2.1.3 A Less Simple Case: Two Overlapping Hypotheses, Equally Likely	9
2.1.4 A Variant of the Less Simple Case: Two Overlapping Hypotheses, with an Initial Bias for One Hypothesis	11
2.1.5 An Even Less Simple Case: Two Overlapping Hypotheses in a Subset Relation, Equally Likely	11
2.1.6 Hypothesis Spaces for Language Learning	16
2.2 Bayesian Updating: General Implementation for Language Learning in a Hypothesis Space with Two Hypotheses	17
2.2.1 Updating with Unambiguous Data in a Hypothesis Space with Two Hypotheses	17
2.2.2 Updating with Ambiguous Data in a Hypothesis Space with Two Hypotheses	21
2.2.3 About t	22
2.3 Summary of Bayesian Updating Adapted to a Linguistic Framework	25
Chapter 3: The Case of Anaphoric <i>One</i>	26
3.1 Anaphoric One: The Necessity of Domain-Specific Constraints	26
3.2 Why Learning Anaphoric One Is Interesting	27
3.3 Anaphoric One	28
3.3.1 Adult Knowledge: Syntactic Categories and Semantic Referents	28
3.3.2 The Pragmatics of Anaphoric <i>One</i>	30
3.3.3 Children’s Knowledge of Anaphoric <i>One</i>	30
3.3.4. Sparse Data for Anaphoric <i>One</i>	31
3.4 Learning Anaphoric One	34

3.4.1	Suggestions for Learning that <i>One</i> is Anaphoric to N'	34
3.4.2	A Regier & Gahl Bayesian Learner	35
3.5	An Equal-Opportunity Bayesian Learner	39
3.5.1	Introducing the Equal-Opportunity Bayesian Learner	39
3.5.2	The Hypothesis Space	40
3.5.3	EO Bayesian Learning	42
3.5.3.1	The R&G Bayesian Learner Implementation	43
3.5.3.2	Updating the Syntax Hypotheses	44
3.5.3.3	Updating the Semantics Hypotheses	48
3.5.4	The Updating Algorithm for Linked Domains	52
3.5.5	What Good Learning Would Look Like	55
3.5.6	Simulating an EO Bayesian Learner	56
3.5.6.2	Semantics	57
3.5.6.3	Linked Domain Updating	58
3.5.6.4	Changing <i>t</i>	58
3.5.7	The Outcome of an EO Bayesian Learner	60
3.6	On the Necessity of Domain-Specific Filters on Data Intake	61
3.7	Deriving the Necessary Domain-Specific Filter	64
3.8	Future Directions	65
3.9	Conclusion	66
Chapter 4:	The Case of Old English Word Order	68
4.1	Filters on Data Intake for Syntactic Learning	68
4.2	Restricting the Data Intake	70
4.2.1	Unambiguous Data	70
4.2.1.1	Unambiguous Data for OV and VO Word Order	70
4.2.2.2	Identifying Unambiguous Data	72
4.2.2.3	Unambiguous Data Summary	74
4.2.3	Simple Clauses	75
4.3	Old English	76
4.3.1	OV and VO word order in Old English	76
4.3.2	Unambiguous Data	77
4.3.2.1	Unambiguous OV	77
4.3.2.2	Unambiguous VO	77
4.3.2.2.1	V2 Interference	77
4.3.2.2.2	Verb-Markers	78
4.3.3	Causes of Language Change	79
4.3.3.1	The Effect of the Unambiguous Data Distribution	79
4.3.3.2	A Concern About Other Causes of Language Change	79
4.3.3.2.1	Scandinavian Influence	80
4.3.3.2.2	Norman Influence	81
4.4	The Model	81
4.4.1	The Individual-Level Model	81
4.4.1.1	Learning in the Individual	81
4.4.1.2	Old English Intake Data	84
4.4.2	Population-Level Model for Old English	85
4.4.2.1	Population-Level Algorithm and Population Growth	85

4.4.2.2 Population Values from Historical Data.....	86
4.4.2.2.1 Procedure to Derive p_{VO} from Historical Data.....	87
4.4.2.2.2 A Concrete Example of Deriving p_{VO} from Historical Data.....	93
4.4.2.3 Answering Questions About Learning Filters	95
4.5 Modeling Results	95
4.5.1 Sufficient Filtering.....	95
4.5.2 Necessary Filtering	96
4.5.2.1 Removing the Unambiguous Data Filter.....	96
4.5.2.2 Removing the Degree-0 Filter	98
4.5.2.3 Removing Both Filters	100
4.6 General Discussion	101
4.6.1 Necessary Filters.....	101
4.6.2 Intake Filtering and Language Change	101
4.6.3 The Feasibility of Filters	102
4.6.4 Future Directions	103
4.6.5 Conclusion.....	104
Chapter 5: The Case of English Metrical Phonology.....	105
5.1 The Unambiguous Data Filter	105
5.1.1 Two Methods For Identifying Unambiguous Data.....	105
5.1.3 Summary	106
5.2 The Cues Method and the Parsing Method.....	107
5.2.1 Cues: Strengths and Weaknesses.....	107
5.2.2 Parsing: Strengths and Weaknesses.....	110
5.2.3 Summary: Cues vs. Parsing Overview.....	111
5.3 The Domain of Metrical Phonology	111
5.3.1 Why English Metrical Phonology?.....	111
5.3.2 Metrical Phonology Parameters.....	112
5.3.2.1 Parameters vs. Probabilistic Association.....	112
5.3.2.2 Parameters for Stress	113
5.3.2.2.1 Quantity Sensitivity	114
5.3.2.2.2 Extrametricality.....	115
5.3.2.2.5 Feet Headedness	119
5.3.2.3 Interacting Parameters.....	120
5.4 The Cues Method for Finding Unambiguous Data.....	121
5.4.1 Quantity Sensitivity.....	121
5.4.2 Extrametricality	122
5.4.3 Feet Directionality	123
5.4.4 Boundedness.....	124
5.4.5 Feet Headedness	125
5.4.6 Summary: Cues.....	126
5.5 The Parsing Method for Finding Unambiguous Data.....	126
5.5.1 The find-all-parses analysis	126
5.5.2 Summary: Parsing.....	127
5.6 Cues and Parsing in a Clean Language Environment: An Easy Case	128
5.7 Learning English: A Harder Case.....	129
5.7.1 Estimating the Composition of the Input to the Learner.....	129

List of Tables

Table 2.1: The effect of data intake accumulation on parameter-setting	24
Table 3.1: The distribution of utterances in the corpus examined by LWF	32
Table 3.2: The expected distribution of utterances in the input to young learners	35
Table 3.3: The expected distribution of utterances in the input to the Bayesian learner for updating the semantics hypothesis.....	50
Table 4.1: OV order advantage in the input for the D0 and D1 clauses.....	85
Table 4.2: Percentage of ambiguous clauses in the historical corpora.....	87
Table 4.3: Formalization of quantities available from historical corpora and quantities to derive	88
Table 4.4: Data quantities after normalization.....	88
Table 4.5: Derived quantities rewritten	90
Table 4.6: Quantities available from historical corpora and quantities to derive	93
Table 4.7: Data quantities after normalization.....	93
Table 4.8: Data from historical corpora and calculate p_{VO}	94
Table 4.9: OV order advantage in the D0 clauses.....	97
Table 4.10: Data gathered from speech directed to young children.....	100
Table 4.11: OV order advantage at 1000 A.D. with no filters.....	100
Table 5.1: A comparison of the properties of the cues and parsing methods	111
Table 5.2: The results of using the cues and parsing methods on Maranungku	129
Table 5.3: Relativize-against-all approach for cues	131
Table 5.4: Relativize-against-potential approach for cues	132
Table 5.5: Relativize-against-potential for parsing	133
Table 5.6: Initial unambiguous data distribution from metrical phonology corpus..	133
Table 5.7: Revised unambiguous data distribution from corpus	134
Table 5.8: Comparison of success of methods and relativization approaches.....	135
Table 5.9: Comparing cues and parsing with different relativization approaches....	140

List of Figures

Figure 1: Two non-overlapping hypotheses, equally probable initially.....	7
Figure 2: Two non-overlapping hypotheses with equal initial probability after seeing various distributions of intake.....	7
Figure 3: Two non-overlapping hypotheses, with an initial bias towards A.....	8
Figure 4: Two non-overlapping hypotheses with an initial bias for A after seeing various distributions and quantities of intake.....	9
Figure 5: Two overlapping hypotheses, with equal probability initially.....	10
Figure 6: Two overlapping hypotheses in a subset relation, with equal probability initially.....	11
Figure 7: Two overlapping hypotheses in a subset relation with equal probability initially, after seeing unambiguous data points for the superset.....	12
Figure 8: Two overlapping hypotheses in a subset relation with equal probability initially, after seeing more subset data points.....	13
Figure 9: Two overlapping hypotheses in a subset relation with equal probability initially, after seeing a mixed distribution.....	14
Figure 10: The Naïve Parameter Learner, learning from subset data only.....	15
Figure 11: Structures for N' strings <i>this ball</i> and <i>this red ball</i>	29
Figure 12: LWF experimental setup.....	30
Figure 13: The set of potential referents for <i>one</i> for an utterance such as <i>I have a red ball, and Jack has one, too</i>	37
Figure 14: The observed set of referents for <i>one</i> for an utterance such as <i>I have a red ball, and Jack has one, too</i>	37
Figure 15: Comparison of different ratios of superset to subset, the likelihood of choosing a member of the subset, and the effect on subset bias.....	38
Figure 16: The syntax hypothesis space.....	40
Figure 17: The semantic hypothesis space.....	41
Figure 18: Linkings across domains.....	42
Figure 19: Double update procedure, part 1.....	53
Figure 20: Double update procedure, part 2.....	54
Figure 21: Type II ambiguous update procedure.....	55
Figure 22: Decision tree for anaphoric <i>one</i> interpretation.....	56
Figure 23: Idealized learner trajectory.....	56
Figure 24: EO Bayesian Learner's trajectory.....	58
Figure 25: Varying <i>t</i>	59
Figure 26: Varying <i>t</i> , with incomplete learning period.....	59
Figure 27: Bayesian learning trajectory with filters.....	64
Figure 28: Cues as underspecified pieces of structure.....	72
Figure 29: Degree-0 unambiguous learning trajectory.....	96
Figure 30: Learning trajectory with ambiguous data.....	97
Figure 31: Learning with degree-1 data.....	99
Figure 32: Metrical phonology parameters.....	113
Figure 33: Sample metrical phonology structure.....	114

