

TO *HAVE* AND TO *BE*: FUNCTION WORD REDUCTION IN CHILD SPEECH,
CHILD DIRECTED SPEECH AND INTER-ADULT SPEECH

by

DANIELLE GILBERTE BARTH

A DISSERTATION

Presented to the Department of Linguistics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2015

DISSERTATION APPROVAL PAGE

Student: Danielle Gilberte Barth

Title: To *HAVE* and to *BE*: Function Word Reduction in Child Speech, Child Directed Speech and Inter-adult Speech

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Vsevolod Kapatsinski	Chairperson
Melissa Redford	Core Member
Tyler Kendall	Core Member
Zhuo Jing-Schmidt	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School
----------------	-----------------------------

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded September 2015

© 2015 Danielle Gilberte Barth
This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs (United States) License.

DISSERTATION ABSTRACT

Danielle Gilberte Barth

Doctor of Philosophy

Department of Linguistics

September 2015

Title: To *HAVE* and to *BE*: Function Word Reduction in Child Speech, Child Directed Speech and Inter-adult Speech

Function words are known to be shorter than content words. I investigate the function words *BE* and *HAVE* (with its content word homonym) and show that more reduction, operationalized as word shortening or contraction, is found in some grammaticalized meanings of these words. The difference between the words' uses cannot be attributed to differences in frequency or semantic weight. Instead I argue that these words are often shortened and reduced when they occur in constructions in which they are highly predictable. This suggests that particular grammaticalized uses of a word are stored with their own exemplar clouds of context-specific phonetic realizations. The phonetics of any instance of a word are then jointly determined by the exemplar cloud for that word and the particular context. A given instance of an auxiliary can be reduced either because it is predictable in the current context or because that use of the auxiliary usually occurs in predictable contexts. The effects cannot be attributed to frequency or semantic weight.

The present study compares function word production in the speech of school-aged children and their caregivers and in inter-adult speech. The effects of predictability in context and average predictability across contexts are replicated across the datasets.

However, I find that as children get older their function words shorten relative to content words, even when controlling for increasing speech rate, showing that as their language experience increases they spend less time where it is not needed for comprehensibility. Caregivers spend less time on function words with older children than younger children, suggesting that they expect function words to be more difficult for younger interlocutors to decode than for older interlocutors. Additionally, while adults use either word shortening or contraction to increase the efficiency of speech, children tend to either use contraction and word shortening or neither until age seven, where they start to use one strategy or the other like adults. Young children with better vocabulary employ an adult-like strategy earlier, suggesting earlier onset of efficient yet effective speech behavior, namely allocating less signal to function words when they are especially easy for the listener to decode.

CURRICULUM VITAE

NAME OF AUTHOR: Danielle Gilberte Barth

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Rice University, Houston, TX

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2015, University of Oregon
Master of Arts, Linguistics, 2009, University of Oregon
Bachelor of Arts, Linguistics, 2006, Rice University

AREAS OF SPECIAL INTEREST:

Corpus linguistics, usage-based linguistics, grammaticalization, information theory, complex predicates, Oceanic languages, language archiving issues

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, University of Oregon, September 2005-Spring 2015

GRANTS, AWARDS, AND HONORS:

Linguistic Society of America Student Abstract Award, 2nd place for *Reduction in Child Speech, Child-Directed Speech and Inter-Adult Speech*: \$300, 2015

Grant from Firebird Foundation for Oral Literature Collection in Matukar, Papua New Guinea: \$9,665, 2013

2nd International Conference on Language Documentation and Conservation student scholarship (for one of 7 top ranked student submissions): \$1000, 2011

Fellowship from Living Tongues Institute for Endangered language and National Geographic Enduring Voices: apx. \$7,000, 2011

Fellowship from Living Tongues Institute for Endangered language and National Geographic Enduring Voices: \$7,500, 2010

Baden-Wuerttemberg-Ministry of Science, Research and the Arts Scholarship for study in Germany, apx. \$4,200, 2009–2010

PUBLICATIONS:

Barth, D. and Kapatsinski, V. (In Press). A multimodel inference approach to categorical variant choice: construction, priming and frequency effects on the choice between full and contracted forms of *am*, *are* and *is*. *Corpus Linguistics and Linguistic Theory*. Published online ahead of Print October 7, 2014: <http://www.degruyter.com/view/j/cllt.ahead-of-print/cllt-2014-0022/cllt-2014-0022.xml?format=INT>

Barth, D. and Anderson, G. D. S. (In Press). Directional constructions in Matukar Panau. *Oceanic Linguistics*.

Anderson, G. D. S., Barth, D. and Rawad Forepiso, K. (2015). The Matukar Panau online talking dictionary: collective elicitation and collaborative documentation, in I Wayan Arka, Ni Luh Nyoman Seri Malini and Ida Ayu Made kam Puspani (Eds), *Language documentation and cultural practices in the Austronesian world: papers from 12-ICAL*, Volume 4. Canberra: Asia-Pacific Linguistics (111-126).

Barth, D., Barth, W. and Rawad Forepiso, K. (2014). *Ngahamam Kabiyai* (Our Customs) DVD. Produced in Eugene, OR.

Barth, D., Barth, W. and Rawad Forepiso, K. (2014). *Ngahamam Neuraurau* (Our Stories) DVD. Produced in Eugene, OR.

Harrison, K. D., Anderson, G. D. S. and Barth, D. (2010-2012). Matukar-English online talking dictionary. <http://matukar.swarthmore.edu/>

Rawad Forepiso, K. and **D. Barth**. (2011). *Learn Panau*. Printed in Matukar: Saky Office.

Rawad Forepiso, K. and **D. Barth**. (2011). *Stories from Matukar*. Printed in Matukar: Saky Office.

Mathieu-Reeves, D. (2005). Direct and Indirect Causation in Sinhala: Examining the Complexity Continuum, in R. Englebretson and C. Genetti (Eds.), *Proceedings from the Workshop on Sinhala Linguistics June 3-4, 2005. Santa Barbara Working Papers in Linguistics*, 17. Santa Barbara, CA: UCSB. http://www.linguistics.ucsb.edu/research/Mathieu-Reeves_vol17.pdf

ACKNOWLEDGMENTS

I sincerely thank Prof. Volya Kapatsinski for being all I could hope for in an advisor and all of his help with this project. Beginning with teaching me advanced statistics over skype in 2010 and getting me to use my first corpus, all the way to today finishing my degree, you have been a wonderful person to work with and I am inspired by you. I also wish to thank Prof. Lisa Redford for letting me use her awesome and unique data and also for all the thought provoking discussions we have had in the last nine years. You have helped me see the forest despite the trees and taught me to be willing to reevaluate theory and see findings in a new light. I wish to thank Prof. Tyler Kendall for helping me get started with the nitty gritty of the corpus project described in this dissertation, introducing me to lots of cool techniques and for letting me use space in his convivial lab. I also wish to thank Prof. Harald Baayen for taking out some time at the Quantitative Investigations in Linguistic Theory 4 to introduce me to ctrees and cforests. Absolutely enormous thanks goes to my wonderful husband, Wolfgang Barth. Not only did you support me emotionally and keep me on schedule, your badass programming skills and help with python scripts, COCA frequency calculations and Buckeye data processing made things go much quicker than they would have otherwise. Thank you, seriously. Finally, I wish to thank audiences at the Linguistics Society of America 2015 Annual Meeting, 8th International Conference on Construction Grammar, Current Trends in Grammaticalization Research and Northwest Phonetics and Phonology Conference and Prof. Zhou Jing-Schmidt, Prof. Rolf Koeppel, and the entire Cognitive Linguistics-Phonology Workgroup at the UO for helpful comments during various stages of this project. All the advice is sincerely appreciated.

To Wolfgang Haus and Jesse Haus

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION.....	1
II. LITERATURE REVIEW.....	9
2.1. Introduction.....	9
2.2. Word Specific Characteristics.....	9
2.2.1. <i>HAVE</i>	10
2.2.1.1. History.....	10
2.2.1.2. Frequency Distribution	12
2.2.2. <i>BE</i>	14
2.2.2.1. History.....	14
2.2.2.2. Frequency Distribution	19
2.3. Factors Influencing Reduction Related to Word Class and Constructions.....	21
2.3.1. Lexical Class and Frequency	21
2.3.2. Construction.....	24
2.3.3. Probability.....	25
2.3.4. Reduction Begets Reduction.....	31
2.4. Factors Influencing Reduction Due to Speaker Characteristics	33
2.4.1. Speech Rate.....	33
2.4.2. Age.....	35
2.4.3. Gender.....	39
2.5. Additional Factors Influencing Reduction.....	41
2.5.1. Preceding Phonological Context.....	41
2.5.2. Priming.....	41
2.5.3. Disfluencies.....	42

Chapter	Page
2.6. Conclusion and Roadmap	42
III. REDFORD CORPUS: REDUCTION IN CHILD AND CHILD-DIRECTED SPEECH.....	45
3.1. Introduction.....	45
3.2. Data	46
3.2.1. Word and Bigram Frequency in a Small Corpus	48
3.2.2. Data Structure Visualization in Redford Corpus	50
3.2.3. Measures of Language Proficiency.....	55
3.3. Methodology	66
3.3.1. Dependent Variables.....	66
3.3.1.1. Normalized Word Duration	66
3.3.1.2. Contraction.....	67
3.3.2. Independent Variables	68
3.3.2.1. Probability and Construction Variables	68
3.3.2.1.1. Probability.....	68
3.3.2.1.2. Construction.....	70
3.3.2.2. Speaker Variables	71
3.3.2.2.1. Age.....	71
3.3.2.2.2. Gender.....	71
3.3.2.2.3. Speaker.....	72
3.3.2.2.4. Syllable and Contraction Rates	73
3.3.2.3. Control Variables	73
3.3.2.3.1. Quartile	73
3.3.2.3.2. Pauses and Disfluencies.....	74

Chapter	Page
3.3.2.3.3. Stress Context	74
3.3.2.3.4. Phonological Context.....	74
3.3.2.3.5. Subject Noun Phrase Type.....	75
3.3.2.3.6. Priming.....	75
3.3.3. Random Forests	77
3.3.4. Multimodel Inferencing	80
3.4. Word Shortening in Child-Directed Speech	82
3.4.1. Introduction.....	82
3.4.2. Random Forest Variable Importance Rankings for Word Shortening in Child-Directed Speech	83
3.4.3. Multimodel Inferencing Analyses for Word Shortening in Child-Directed Speech	89
3.4.4. Conclusion for Word Shortening in Child-Directed Speech	100
3.5. Word Shortening in Child Speech	102
3.5.1. Introduction.....	102
3.5.2. Random Forest Variable Importance Rankings for Word Shortening in Child Speech	103
3.5.3. Multimodel Inferencing Analyses for Word Shortening in Child Speech....	113
3.5.4. Conclusion for Word Shortening in Child Speech	121
3.6. Contraction in Child-Directed Speech	124
3.6.1. CDS Contraction of <i>HAVE</i>	124
3.6.2. Contraction of <i>BE</i>	125
3.6.2.1. Introduction.....	125
3.6.2.2. Random Forest Variable Importance Rankings.....	126
3.6.2.3. Multimodel Inferencing	128
3.6.2.4. Conclusion for BE Contraction in Child-Directed Speech	131

Chapter	Page
3.7. Contraction in Child Speech	132
3.7.1. CS Contraction of <i>HAVE</i>	132
3.7.2. CS Contraction of <i>BE</i>	133
3.7.2.1. Introduction.....	133
3.7.2.2. Random Forest Variable Importance Ranking	134
3.7.2.3. Multimodel Inferencing	140
3.7.2.4. Conclusion for <i>BE</i> Contraction in Child-Speech.....	144
3.8. Discussion and Conclusion	145
3.8.1. Results Summary	146
3.8.2. Information Compression	148
3.8.3. Child Language Proficiency	151
3.8.4. Child Directed Speech Features for 5-10 Year-olds.....	157
3.8.5. Methods Comparison.....	158
IV. REDUCTION IN INTER-ADULT SPEECH	164
4.1. Introduction.....	164
4.2. Data and Methodology.....	164
4.2.1. Data.....	164
4.2.2. Dependent Variables.....	165
4.2.2.1. Normalized Duration	165
4.2.2.2. Contraction.....	165
4.2.3. Independent Variables	165
4.2.3.1. Probability and Construction Variables	165
4.2.3.1.1. Probability.....	165

Chapter	Page
4.2.3.1.2. Construction.....	166
4.2.3.2. Speaker Variables	166
4.2.3.2.1. Age and Gender	166
4.2.3.2.2. Speaker.....	167
4.2.3.2.3. Syllable and Contraction Rates.....	167
4.2.3.3. Control Variables.....	168
4.2.3.3.1. Quartile	168
4.2.3.3.2. Stress Context	168
4.2.3.3.3. Phonological Context.....	168
4.2.3.3.4. Subject Noun Phrase.....	168
4.2.3.3.5. Priming.....	169
4.2.4. Statistical Procedure.....	169
4.3. Results.....	170
4.3.1. Duration	170
4.3.1.1. Word Distribution in Sample.....	170
4.3.1.2. Random Forests	170
4.3.1.3. Multimodel Inferencing Full Model	174
4.3.1.4. Multimodel Inferencing – Word Specific Results: <i>BE</i>	182
4.3.1.5. Multimodel Inferencing – Word Specific Results: <i>HAVE</i>	190
4.3.2. Contraction.....	196
4.3.2.1. Word Distribution in Sample.....	196
4.3.2.2. Random Forests	198
4.3.2.3. Multimodel Inferencing – Word Specific Results: <i>BE</i>	204

Chapter	Page
4.3.2.3.1. All <i>BE</i> Data Collapsed	204
4.3.2.3.2. <i>is- 's</i>	208
4.3.2.3.3. <i>are- 're</i>	209
4.3.2.3.4. <i>am- 'm</i>	211
4.3.2.3.5. Conclusion	214
4.3.2.4. Multimodel Inferencing – Word Specific Results: <i>HAVE</i>	214
4.3.2.4.1. All <i>HAVE</i> Data Collapsed.....	214
4.3.2.4.2. <i>have- 've</i>	218
4.3.2.4.3. <i>has- 's</i> and <i>had- 'd</i>	221
4.3.2.4.4. Conclusion	221
4.4. Discussion and Conclusion	222
4.4.1. Corpus Comparison: Redford Corpus vs. Buckeye Corpus.....	224
4.4.2. Construction Effects.....	225
4.4.3. <i>BE</i> – Contraction vs. <i>HAVE</i> – Duration Shortening	232
V. CONCLUSION	234
5.1. Introduction.....	234
5.2. Child and Adult Reduction Strategies.....	236
5.3. Child Reduction Strategies Over Time	236
5.4. Caregiver v. Inter-Adult Speech Reduction Strategies	238
5.5. Usage-Based Speech Production	239
5.6. Grammaticalization.....	246
REFERENCES CITED.....	249

LIST OF FIGURES

Figure	Page
1. COCA <i>HAVE</i> Frequency Distributions	13
2. Buckeye Corpus <i>HAVE</i> Frequency Distributions	14
3. Redford Corpus <i>HAVE</i> Frequency Distributions	14
4. COCA <i>BE</i> Frequency Distributions.....	19
5. Buckeye Corpus <i>BE</i> Frequency Distributions	20
6. Redford Corpus <i>BE</i> Frequency Distributions	20
7. Numbers of stories by age of children in months	48
8. Structure of the Child Speech and Child-Directed Speech Corpus.	51
9. Structure of the Child Speech Texts	52
10. Structure of the Child-Directed Speech Texts	53
11. Structure of one Child Speech text	53
12. Structure of one Child-Directed Speech text	54
13. Text Entropy by Age in Children’s Speech	58
14. Text Entropy by Age in Caregivers’ Speech	59
15. Syllable Rate by Child Age in Caregiver’s Speech	60
16. Syllable Rate by Child Age in Children’s Speech	61
17. Syllable Rate by Entropy in Children’s Speech.....	62
18. Syllable Rate by Entropy in Caregiver’s Speech.....	63
19. Caregiver-Child Dyad Text Entropy Correlation	64
20. Caregiver-Child Dyad Syllable Rate Correlation	65
21. Example Classification Tree	78
22. Example Classification Tree on New Data Subset.	79
23. Speaker-specific and word-specific variable importance for CDS grammatical word duration	85
24. Preceding Context variable importance for CDS grammatical word duration	87
25. Following Context variable importance for CDS grammatical word duration.....	89
26. Interaction of Speaker Contraction Rate and Speaker Syllable Rate on CDS word duration	96
27. Interaction of Speaker Contraction Rate and Child Age on CDS word duration	97

Figure	Page
28. Partitioning of Speaker Contraction Rate on CDS word duration	98
29. Speaker-specific and word-specific variable importance for CS grammatical word duration	104
30. Speaker-specific and word-specific variable importance for CS grammatical word duration by school grade	107
31. Preceding Context variable importance for CS grammatical word duration	108
32. Preceding Context variable importance for CS grammatical word duration by school grade	109
33. Following Context variable importance for CS grammatical word duration	111
34. Following Context variable importance for CS grammatical word duration by school grade	112
35. Interaction of Speaker Contraction Rate and Speaker Syllable Rate on CS word duration	117
36. Interaction of Speaker Contraction Rate and Speaker Age on CS word duration	118
37. Correlation of Caregiver and Child Dyad Contraction Rates	119
38. Child Normalized Duration by Text Entropy and Age	123
39. CDS Contraction rate of Perfect <i>HAVE</i> by Speaker	125
40. CDS Contraction rate of Perfect <i>BE</i> by Speaker	125
41. CDS Word and Speaker-specific factor variable importance ranking for contraction	126
42. CDS preceding context factor variable importance ranking for contraction	127
43. CDS following context factor variable importance ranking for contraction	129
44. CS Contraction rate of Perfect <i>HAVE</i> by Speaker	133
45. CS Contraction rate of Perfect <i>BE</i> by Speaker	134
46. CS word and speaker specific factors variable importance ranking for contraction	135
47. CS word and speaker specific factors variable importance ranking for contraction by grade	136
48. CS preceding context variable importance ranking of contraction	137
49. CS preceding context variable importance ranking of contraction by grade	138
50. CS following context variable importance ranking of contraction	139
51. CS following context variable importance ranking of contraction by grade	140

Figure	Page
52. CS Contraction by Quartile.....	143
53. Caregiver contraction rates by speaker	149
54. Child contraction rates over time by speaker.....	150
55. 7+ Child contraction rates over time by speaker	151
56. Relationship of Speaker Age and Syllable Rate on Duration Reduction in Child Speech.....	152
57. Relationship of Speaker Age and Text Entropy on Duration Reduction in Child Speech.....	154
58. Relationship of Contraction Rate and Text Entropy on Duration Reduction in Child Speech.....	155
59. Relationship of Contraction Rate and Syllable Rate on Duration Reduction in Child Speech.....	156
60. Schema of caregiver content to function word duration proportion.....	158
61. Random forest importance ranking of speaker-based and word-based factors for normalized word duration in IAS	172
62. Random forest importance ranking of preceding context factors for normalized word duration in IAS	173
63. Random forest importance ranking of following context factors for normalized word duration in IAS	174
64. Following Joint Probability and Transitional Probability on Duration for IAS	179
65. Preceding Joint Probability and Transitional Probability on Duration for IAS.....	180
66. Distribution of <i>was</i> phonemic transcriptions in the Buckeye Corpus.....	183
67. Distribution of <i>is</i> phonemic transcriptions in the Buckeye Corpus	185
68. Distribution of <i>are</i> phonemic transcriptions in the Buckeye Corpus	186
69. Distribution of <i>were</i> phonemic transcriptions in the Buckeye Corpus.....	187
70. Distribution of <i>had</i> phonemic transcriptions in the Buckeye Corpus.....	191
71. Distribution of <i>has</i> phonemic transcriptions in the Buckeye Corpus	192
72. Distribution of <i>have</i> phonemic transcriptions in the Buckeye Corpus	193
73. IAS Contraction rate of Perfect <i>HAVE</i> by Speaker	197
74. IAS Contraction rate of <i>BE</i> by Speaker	199
75. Random forest importance ranking of preceding context factors for contraction in IAS.....	200

Figure	Page
76. IAS Contraction by Utterance Quartile.....	201
77. Random forest importance ranking of preceding context factors for contraction in IAS	202
78. IAS Contraction Rate by NP Type for <i>BE</i>	203
79. Random forest importance ranking of following context factors for contraction in IAS	204
80. IAS Contraction of <i>AM</i> by Utterance Quartile	213
81. IAS Contraction of <i>HAVE</i> by Inflection	216
82. IAS Contraction of <i>have~'ve</i> by Subject Noun Phrase Type	217
83. Following TP by Construction Type for <i>BE</i>	227
84. Following TP by Construction Type for <i>HAVE</i>	228

LIST OF TABLES

Table	Page
1. Ten most frequent words and bigrams in the COCA and in the Redford Corpus	49
2. Ranges of Normalized Durations by Speaker Group.....	67
3. Proportion of contraction for <i>is, am, are, have, had</i> and <i>has</i> by corpus	68
4. Pearson Correlation Matrix of CDS Following Context Variables for CDS Duration Analysis	91
5. CDS Duration Multimodel Inferencing Results	92
6. Models of CDS duration with a Δ below 2	93
7. CDS <i>HAVE</i> Duration Multimodel Inferencing Results	99
8. Models of CDS <i>HAVE</i> duration with a Δ below 2.....	99
9. Child Speech Duration Multimodel Inferencing Results.....	116
10. Models of Child Speech duration with a Δ below 2.	117
11. CS <i>HAVE</i> Duration Multimodel Inferencing Results	120
12. Models of CS <i>HAVE</i> duration with a Δ below 2.....	121
13. Multimodel Inferencing Output of CDS Contraction	130
14. Models of CDS Contraction with a Δ below 2.....	131
15. Multimodel Inferencing Output of CS Contraction	142
16. Models of CS Contraction with a Δ below 2.	144
17. Redford Corpus Results by Dependent and Independent Variables	147
18. Bigram probability examples from the Buckeye Corpus.....	166
19. Word distribution and length ranges.....	170
20. IAS Duration Multimodel Inferencing Results.....	176
21. Models of IAS duration with a Δ below 2	177
22. IAS Duration Multimodel Inferencing Results: <i>BE</i>	188
23. Models of IAS duration with a Δ below 2	189
24. IAS Duration Multimodel Inferencing Results for <i>HAVE</i>	194
25. Models of IAS duration with a Δ below 2 for <i>HAVE</i>	195
26. Contraction by Lemma in the Buckeye Corpus	197
27. IAS Contraction of <i>BE</i> Multimodel Inferencing Results	207
28. Models of IAS contraction of <i>BE</i> with a Δ below 2.....	208

Table	Page
29. IAS Contraction of <i>IS</i> Multimodel Inferencing Results	210
30. Models of IAS contraction of <i>IS</i> with a Δ below 2	211
31. IAS Contraction of <i>AM</i> Multimodel Inferencing Results	211
32. Models of IAS contraction of <i>AM</i> with a Δ below 2.....	212
33. IAS Contraction of <i>HAVE</i> , <i>HAD</i> and <i>HAS</i> Multimodel Inferencing Results	219
34. Models of IAS contraction of <i>HAVE</i> , <i>HAD</i> and <i>HAS</i> with a Δ below 2.....	220
35. IAS Contraction of <i>HAVE</i> Multimodel Inferencing Results.....	220
36. Models of IAS contraction of <i>HAVE</i> with a Δ below 2	220
37. Buckeye Corpus Results by Dependent and Independent Variables	223
38. Mean probability contexts by Construction	226

CHAPTER I

INTRODUCTION

The English words *BE* and *HAVE* are old and complex. Both words have multiple meanings as part of different constructions and vary in degree of grammaticalization. In its possessive use (*I have a dog*), *HAVE* is usually considered to be a content word, while in auxiliary uses (*have written*) it is a function word, with semi-modal uses (*have to*) in between. *BE* is somewhat more grammaticalized in its main (copula) verb usage (*Dachshunds are long*), where it is considered grammatical despite being a main verb. It is also a function word in all of its auxiliary usages (e.g. *am writing*). As they vary in where they are on the content/function cline, the different uses of *BE* and *HAVE* provide a useful window on form variation among grammaticalized constructions.

The different *BE* and *HAVE* constructions have different levels of usage frequency. Interestingly, as detailed in Chapter II, the less grammaticalized uses of the two words are the most frequent ones, which is not (thought to be) the usual case since function words are prototypically more frequent than content words (Bybee, 2007, 2011).

These words therefore allow us to deconfound frequency and other factors in the phonetic and phonological reduction that usually accompanies grammaticalization. If there is indeed a difference in level of reduction due to meaning, we would expect the more grammaticalized uses of *HAVE* and *BE* (i.e. auxiliaries) to have the highest level of reduction. Since the grammaticalized meanings of *BE* and *HAVE* are less frequent than their source meanings (possessive and copular), greater levels of reduction in the more

grammaticalized auxiliary uses could not be attributed to frequency (cf. Bybee 2011, Traugott 2011). Instead, the greater reduction could be attributed to lower semantic weight of the more grammaticalized uses (Bybee and Pagliuca, 1985; Bybee, Perkins and Pagliuca, 1994; Gabelentz, 1891; Givón, 1985; Heine, 1993; 2003; Hopper and Traugott, 1993; Lehmann, 1995). In information-theoretic terms, lower semantic weight decreases information content but so does greater predictability in context: what is expected is not informative and therefore also not surprising. Langacker (2011) argues that accompanying grammaticalization is diminished attention and salience (as one would expect of relatively uninformative elements), and that this diminished attention and salience results in signal compression, where duration, amplitude and pitch level are reduced, which in turn results in diminished attention and salience. In the present dissertation, I show that average predictability of a word use in a constructional context is indeed associated with reduction, as it is only some grammaticalized uses of a word (namely, ones that tend to be predictable in context) that are associated with greater reduction relative to the source construction.

In this dissertation, I attempt a usage-based description of reduction patterns in *BE* and *HAVE* in the speech behavior of adults speaking to other adults, adults speaking to children and children speaking to adults. By “description” I mean a fairly exhaustive investigation of the factors influencing reduction of *BE* and *HAVE*, focusing particularly on the roles of probability and syntactic construction, as these predictors are of fundamental theoretical importance in the usage-based framework I follow here.

By ‘reduction’ I mean adjustments to form that happen when that form conveys little information, i.e., when it is predictable in the context. These adjustments most

prototypically affect duration (e.g. Baker and Bradlow, 2009) but may also alter the spectral characteristics of speech (e.g. Aylett and Turk, 2004; Wright, 2004). The present dissertation adopts an exemplar-based framework for thinking about how this kind of form variation is implemented: reduced forms of various kinds are selected from a richly specified lexicon of articulatory exemplars tagged with various contextual characteristics (e.g. Pierrehumbert, 2001; 2002). Framing reduction as selection of ‘reduced’ exemplars, associated with low-informativity contexts, allows for investigating both ‘phonetic’ reduction (shortening) and ‘phonological’ reduction (contraction) in a unified framework (as opposed to treating the former as a matter of phonetic implementation and the latter as an outcome of a variable rule à la Labov, 1969). In the exemplar-based framework, both contraction and shortening reflect selection of a context-appropriate exemplar in low-informativity contexts. In some forms of *HAVE* and *BE*, shortening and contraction are alternatives that occur in similar contexts and achieve the same goal, although there are also contexts where only shortening is available.

Note that by adopting this selection framework I do not wish to deny that there are also online parametric adjustments to, particularly, duration and fundamental frequency (F0), driven by factors like gender, speech rate (Dilts, 2013), disfluency for buying time for lexical access (Schnadt, 2009) and, as I argue below, individual’s reduction strategy preferences. These factors do not affect contraction but do affect shortening. Therefore these factors are also included in this dissertation alongside probability. Other factors known to influence duration and F0 in specific contexts, like contrastive focus on function words (Mathieu-Reeves and Redford, 2009), are not included because of their rarity in casual speech.

Constructions are form-meaning pairings, and they are thought to be the basic units of language by proponents of usage-based linguistics and usage-based construction grammar in particular (Goldberg, 1995; 2006). The same form, like *is* or *have*, is not always an instance of the same word-level construction, and different word-level constructions tend to combine with different phrasal and sentence-level constructions. One aim of the present dissertation is then to see whether reduction patterns differ by construction. The word-specific phonetics approach of Pierrehumbert (2002) is extended here to grammaticalized words in the context of phrase-level construction such as the English Perfect. The language learner is thought to learn and remember richly specified exemplars of words tagged with attributes of contexts in which the exemplars occurred, including the semantics that identify the construction containing the word. However, when the learner aims to produce the word in a certain construction, they may at least occasionally select an exemplar that is not particularly suited to the present context but is highly representative of previous pronunciations of that word in that construction: the context and the word-in-construction cloud of exemplars are jointly determining the produced phonetics. Thus, for example, a word-level construction that tends to be probable in context will become associated with many reduced exemplars. A speaker will then be likely to produce a reduced exemplar of the word even when the context calls for an unreduced exemplar (Bybee, 2002; Bybee and Torres Cacoullos, 2008; Raymond and Brown, 2012). As a result, words that tend to occur in different contexts may gradually drift away from each other, acquiring distinct pronunciations (Bybee, 2001). At the phonological contraction level, this is true for *HAVE* in American English, where only

the perfect *HAVE* is eligible for contraction while the semi-modals can fuse with the following *to* (*hafta, hasta, hadda*), and the source possessive *HAVE* can do neither.

I look at the behavior of *HAVE* and *BE* in the speech of school-aged children, their caregivers and in inter-adult speech. We know that adult speakers' productions of function words are generally reduced (shorter, less prominent, cliticized, etc.) in comparison to their production of content words. This is the case when parents are speaking to their children (Brown, 1973; Swanson, Leonard and Grandour, 1992) resulting in children acquiring first content words, then function words, despite the much higher frequency of function words in running speech. However, by age five, children have acquired function words and content words alike in that they generally produce them in the correct contexts. This study examines *HAVE* and *BE* in child speech, in addition to adult speech, to assess whether the words' productions are also affected by construction and probability in the same way for both speaker populations. Function words are susceptible to reduction in adult speech because of their frequency, their predictability in context, their low informativity, their low semantic weight and because they tend to be unstressed. However, children are less proficient at producing reduced syllables, so their productions could potentially be less affected by these factors. If children's productions are affected by frequency, probability and informativity, then they are tracking either specific statistical information about word co-occurrences or generalizations derived from those statistics, just like adults. Furthermore, if the influences of predictability-related variables increase with age, then children must not simply get better at producing weak syllables as they get older; rather, they must be getting a better handle on contextual probabilities and/or improve in being able to utilize

the learned probabilities in online language production. Better production of weak or unstressed syllables is a sign of improved speech production, but improved tracking and using statistical information about word frequency, predictability and informativity is a sign of improved language production.

Given the focus of the present work on describing linguistic behavior, I adopt an information-theoretic approach to statistical analysis (Burnham and Anderson, 2002). Instead of being concerned with statistical significance (i.e., the likelihood of obtaining the observed data if the predictor had no effect), the information-theoretic approach is concerned with predictiveness. In the present case, we are interested in the variables that will help us predict characteristics of reduction behavior in (future) samples of speech from the same kinds of speakers. The analyses below focus on a measure of variable predictiveness called ‘cumulative probability’. When cumulative probability is above .5, models that include the variable in question are more predictive (are better models) than models that exclude it. All such variables are therefore considered as useful for describing the observed reduction behavior.

The problem with focusing on significance in an exploratory study is that it discourages careful examination of the data (Kruschke, 2010): the fewer predictors one examines, and the fewer ways one looks at the data, the easier it is to achieve significance (because of the ‘multiple comparisons problem’). Indeed, if one were to, say, only look at any of the probabilistic predictors in the present study, it would be massively significant, while if one were to examine all predictors we know to play a role in reduction, none would be significant in any reasonably-sized dataset. The information-theoretic approach

tries to avoid this problem by focusing on whether the predictors overall improve our ability to predict (Burnham and Anderson, 2002).

The current work contributes to our knowledge of speech production by combining strains of research on word reduction in speech production, grammaticalization and child language acquisition. Chapter II presents a literature review summarizing relevant theories of speech production and the factors that influence word reduction, particularly for function words, before turning to a summary of additional factors relevant to child and caregiver productions of function words and then finally discussing the words *BE* and *HAVE*, summarizing their history in English, their acquisition by children and the frequency distributions of their meanings (copular, future, passive and progressive and modal, perfect and possessive, respectively). The next three chapters present the results from two corpus studies and one experimental study investigating function word production. Chapter III presents results from the Redford Corpus, a corpus developed from picture book narrations of school-aged children and their caregivers. As detailed in chapter III, the data was collected by Prof. Melissa A. Redford as part of her larger study on the development of prosody in the speech of school-aged children. Chapter III is the longest chapter of this dissertation because this is the first detailed, quantitative investigation of the production of function words by this age group or their caregivers. I conclude that children must develop strategies for reduction as part of their acquisition of language and that as they get older they learn what words are uninformative as part of learning about word meanings. As they learn more about the unimportance of function words, they are able to speak more efficiently and in a more adult-like manner. Chapter IV presents results from the Buckeye Corpus

(Pitt, et al, 2007), a corpus of interviews of adults from Ohio. The Buckeye Corpus is much larger than the Redford Corpus, allowing us to confirm results from the smaller corpus, as well as investigate some word inflection-specific effects. Finally chapter V offers a discussion of the results and their place in the literature on speech production and grammaticalization theory.

CHAPTER II

LITERATURE REVIEW

2.1. Introduction

In this chapter, I first summarize the etymological history and frequency distribution of the function words under examination in this study: *BE* and *HAVE*.¹ Next, I describe the importance of lexical class, frequency and predictability in predicting reduction, including a discussion of theories that account for the occurrence of probabilistic reduction. I then go on to describe speaker-specific characteristics that are known to influence reduction, including factors that are important to consider when examining child speech. Finally, I outline other factors that are necessary to take into account when examining patterns of reduction in both adult and child speech.

2.2. Word Specific Characteristics

The specific words under investigation in this study are inflections of *BE*: *am*, *are*, *is*, *was*, *were* and inflections of *HAVE*: *had*, *has*, *have* (capitalized words indicate a reference to all word inflections, lower case words indicate a reference to a particular word form). These particular inflections have been selected because they are frequent enough in short texts that there is a reasonable amount of data to investigate. *HAVE* and *BE* have different meanings that vary in frequency which allows for an empirical investigation if word meaning matters for word form reduction and if so, if it is the form with the most frequent or the most grammaticalized word meaning that is the most

¹ Capitalized words represent a lemma: word and all its inflections. I use the capitalized words *BE* and *HAVE* to stand for all meanings of these words as well.

reduced. In this section I present a brief history of *HAVE* and *BE*, showing which uses are more grammaticalized. I also examine the meanings' frequency distributions in several corpora, showing that usage frequency is not aligned with degree of grammaticalization in this case.

2.2.1. *HAVE*

2.2.1.1. *History*

The possessive construction is the source for the perfect and modal constructions. It is the oldest *HAVE* construction and was present in Old English. The beginnings of the perfect construction also developed during the Old English period. In Old English, a construction consisting of a possessive verb followed by a past participle was still considered a possessive construction and usually had an object explicitly expressed, with the past participle agreeing with the object in gender and number (Visser, 1973: 2189). Over time, but still in the Old English period, there developed ambiguity between this construction expressing a state that had arisen from an antecedent action or expressing a completed action, as in (1) below (ibid).

1) *gyt ge habbaþ eowre heortan geblende?*

have ye your heart hardened? (Visser, 1973: 2189 citing the Old English Gospels)

Although possessives and perfects are not ambiguous in Present-Day English due to word order (cf. *I have my work done* v. *I have done my work*) this word order difference did not disambiguate resultative antecedent action (possessive construction) and completed action (perfect construction) in Old English, allowing for semantic ambiguity between the two (Visser, 1973: 2189). Later in Old English, this construction could be used with transitive verbs without a direct object, leading to even more

ambiguity between the expression of resultant states and completed actions (Visser, 1973: 2191). By Middle English, the construction could use intransitive verbs as well as transitive verbs (Visser 1973: 2191-2192) as in *They have gone* and therefore we can clearly state that a new construction, the perfect, existed as there is no ambiguity because an intransitive (de facto no object), such as *gone*, does not describe a possessed object in a resultative state; it clearly describes a completed action.

The past tense of the perfect, the pluperfect, also began its development in Old English with the same trajectory: *HAVE* + object + past participle of transitive verb > *HAVE* + past participle of transitive verb > *HAVE* + past participle of intransitive verb (Visser, 1973: 2212).

Modal *HAVE* developed from possession to meaning possession plus the ability to do something (Visser, 1969: 1474), as in (2) below during Old English. Although (2) is a Middle English example, similar constructions were found in Old English. Later in Old English, the possessive meaning weakened and the construction gained an additional meaning of duty and obligation, as in (3) below (a Modern English example, but similar examples were present in Old English). Finally, around the end of the Old English period, *HAVE* appears directly before *to* and the infinitive verb and clearly indicates obligation and not possession. Any object in the clause is an object of the infinitive now, rather than *HAVE*, meaning *HAVE* is clearly a function word (Visser, 1969: 1478, and cf. van der Gaaf, 1931), as in (4) below (a Modern English example but other such examples were found in Middle English).

2) *they hadden na more to giuen*

‘They had no more to give / they could not give anymore’ (Visser, 1969: 1476).

3) *He hath some message to deliver us*

(Visser, 1969: 1477 citing Shakespeare (1588) *Titus IV*).

4) *It's getting dark; so I have to stop now*

(Visser, 1969: 1478)

This last use was rare in Old and Middle English and did not firmly become established until Modern English (Visser, 1969). Some researchers do not consider these kinds of instances to actually be modal auxiliaries in Old English (Brinton, 1991; Bock, 1931; Fischer, 1994; Mitchell, 1985). Fischer (1994) argues that the semantic ambiguity in meanings is not enough to clearly say that a modal usage has developed and argues instead for a structural criterion (word order change) to determine the onset of new construction, which does not occur until late Middle English.

2.2.1.2. *Frequency Distribution*

The Corpus of Contemporary American English (COCA) (Davies, 2008-) is a large corpus of 450 million words consisting of newspapers, fiction, magazines, academic publications and a spoken section of transcripts of radio and television news programs. The spoken section consists of 95 million words. Figure 1 shows the distribution of the three *HAVE* construction types in the spoken section of the COCA. *Have* is clearly the most common inflection, likely because it is the word form used for both the third person plural and the infinitive. The possessive and perfect constructions occur at relatively equal frequencies, but the modal construction is far less frequent. These patterns manifest themselves again in the Buckeye Corpus (Pitt et al., 2007), a smaller 307,000 word corpus of adult interviews. But the patterns are on a much smaller scale, as shown in

Figure 2, although there is proportionally more contraction in the Buckeye Corpus.

Figure 3 shows the patterns of *HAVE* frequency distribution in the Redford Corpus, a 78,000 word corpus of child and caregiver story narrations. For this corpus data, the distribution patterns are slightly different. The word form that occurs most often is *had*, because most of the story telling done by caregivers and children is done in the past tense. There are also far fewer perfect constructions in the Redford Corpus, although modal constructions are still the least frequent construction type. Additionally, there is proportionally less perfect construction contraction than in either of the two larger corpora.

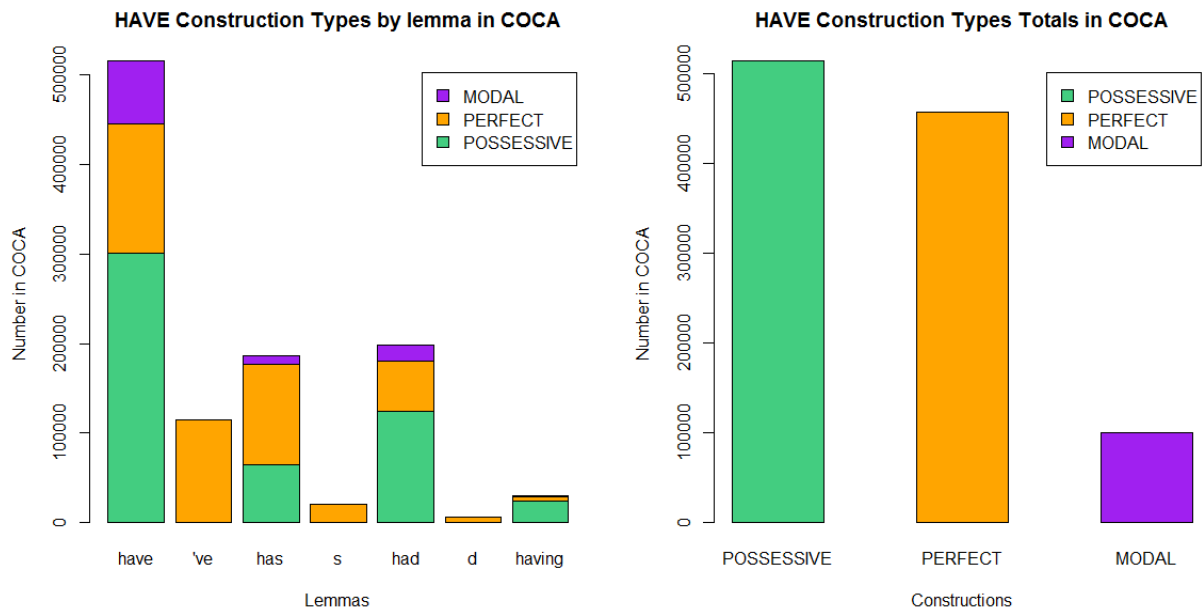


Figure 1. COCA *HAVE* Frequency Distributions

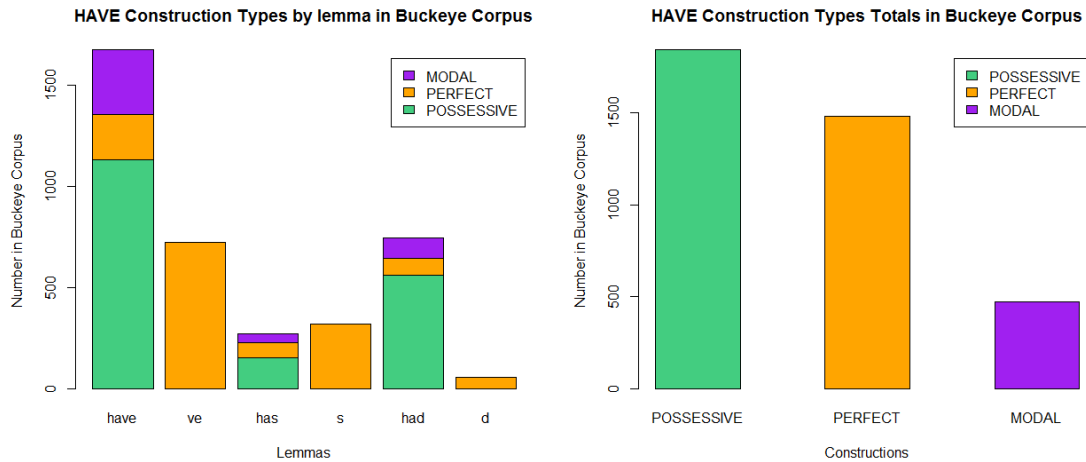


Figure 2. Buckeye Corpus *HAVE* Frequency Distributions

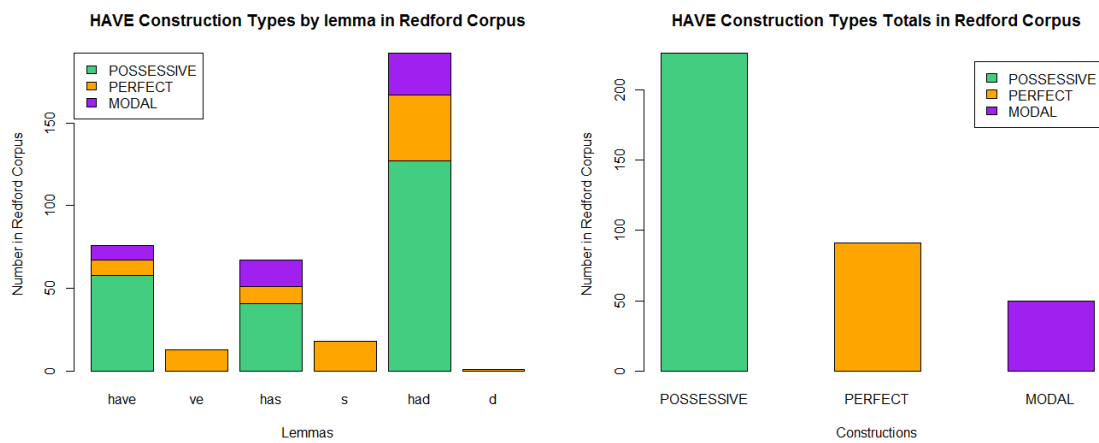


Figure 3. Redford Corpus *HAVE* Frequency Distributions

2.2.2. *BE*

2.2.2.1. *History*

Copula constructions with *BE*, as well as the beginnings of the progressive and passive constructions with *BE*, were present in Old English. The copula construction had an inflection of *BE* as the main verb and a predicate. Example (5) shows this, with *beo*

'am' and a prepositional phrase beginning with *mid* 'with.' The copula construction with *BE* has not changed greatly since then. The main verb has changed and limited its inflections, but it is still found in the same syntactic position and takes the same kind of predicates, i.e., prepositional, adjectival and nominal.

5) *Ic beo mid eow ealle dagas*

'I am with you always' Gospel Matthew 28:20 cited by Visser (1963:160)

There is debate as to the source construction for the progressive construction but both potential sources are copula constructions. The first is adjectival: a construction of *BE* with a present participle, i.e. *he wæs huntende*. The second is locative: a construction of *BE* with a preposition and a gerund, i.e. *he wæs on huntunge* (Leech et al. 2009:120).

The first potential progressive source is a durative aspect construction found in Old English, as in (6-7), which sometimes was indistinguishable from simple tense, as in (8). It could be used to express an ingressive aspect as well (9) (Mitchell and Robinson 2001:110, Quirk and Wrenn 1957:79-80). The present particle has the inflectional ending <ende>. However, durative aspect could alternately be expressed without using this kind of construction, using another structure instead.

6) *ic mē gebidde to ðāem Gode þe bīō eardigende on heofonum*

'I pray (at this moment) to the God who is dwelling (not only at this moment) in the heavens' (Quirk and Wrenn 1957:80).

7) *Ðær wæron sume of ðām bōcerum sittende, and on hiera heortum ðencende*
(*Erant...sedentes...cogitantes*)

'There were some of the scribes sitting there and thinking in their hearts' (Quirk and Wrenn 1957:80).

8) *Ða wæs se cyning openlice andettende Ðam biscope*

'The king openly confessed to the bishop' (Mitchell and Robinson 2001:110)

9) *þætte nāenig...wære āwendende þās ūre dōmas*

'that no one should set about changing these our decrees' (Quirk and Wrenn 1957:80).

In Middle English, one could still express the durative or progressive aspect without using the progressive construction, but there was a dramatic increase in its frequency for expressing these aspects. During the Middle English period, the present participle started to occur with the ending <ung> or <ing>. Jones (1972) claims this is likely due to analogy with the deverbial noun or gerund that ended with a vowel followed by <ng(e)>. This ending replaced <ende> in many contexts, including in the progressive construction (Visser 1966:1089-94). The variation between these two ending forms in the progressive construction starts around 1250.

The locative type construction, as in 'he is on huntung', is the second possible source for the progressive construction. In this scenario, the source for the progressive was a copula followed by a preposition and then the present participle with a <ung> or <ing> ending, not the <ende> ending. It is also quite possible that the progressive

construction developed under the simultaneous influences of both of these constructions. In any case, the <ende> type participles were obsolete after 1500, and the ‘on hunting/a hunting’ became dialectal in Modern English, leaving the <ing> participles, without prepositions, to become the present day form of the progressive construction (Visser 1966:1094-5). Typical examples of the Middle English progressive can be seen in (10-11).

10) *Oððo swa hwar swa heo sy sittende, standende, oððo gangende, æfre beo hniwiende mid hyre heafede*

‘Wherever he may be, sitting, standing or walking, let him always be with head bowed down’ Rule of St. Benedict cited by Visser (1966:1095)

11) *Heo...iuunden Þene king Þær he wes an slæting*

‘and they found the king where he was hunting’ Layamon’s Brut cited by Visser (1966:1095)

The progressive doubled in use between 1500 and 1570 and then doubled again between 1640 and 1710. The rate of progressive use has also increased by 45% from 1950 to the present in spoken British English, based on searches in the Diachronic Corpus of Present-Day Spoken English (Leech et al. 2009:121, 126). Progressive constructions are becoming more and more frequent as they are replacing simple present tense to describe ongoing actions.

There were a few ways of expressing the passive in Old English, including a passive construction with *BE* (12). There was also another type of passive construction with the copula

'become' *woerðan*. The former was typically used in durative constructions, the later in perfective, but with lots of variation between the two. Authors tended to use just one or the other type. The form of this construction was an inflection of *BE* and a past participle (Quirk and Wrenn 1957:80-81).

12) *Ne bið ðær nāenig ealo gebrowen*

'No ale is (ever) brewed there' (Quirk and Wrenn 1957:80)

In Middle English most passive constructions occurred with the *BE* construction (13), but could still be expressed with the copula 'become', (at that time *worte*) (Burrow and Turville-Petre 1996:52).

13) *he...wæs wæl underfangen fram ðe pape Eugenie*

'He was well received by Pope Eugenius' (Burrow and Turville-Petre 1996:52)

In present day English, there is a decrease in the use of the passive construction with *BE*, as it is being replaced with the *got* passive construction (*he got picked for the team*). There were 90 instances of the *BE* past participle passive per 10,000 words in 1650, but only 30/10,000 words in 1990 based on searches in A Representative Corpus of Historical English Registers (ARCHER) (Hundt 2004:116). In spoken American English, there is a 28.2% decrease in the use of the passive from 1950 to 1990 based on searches

in the Brown Family of Corpora (Leech et al. 2009:148).

2.2.2.2. Frequency Distribution

The copula is an extremely frequent word, with *is* always being one of the top-ten words in corpora of running speech. Passive constructions are infrequent in inter-adult speech (Xiao, McEnery and Qian, 2006) child-directed speech (Gordon and Chafetz, 1991) and in child speech (Pinker, Lebeaux and Frost, 1987; Israel, Johnson and Brooks, 2000). The frequency of progressive constructions is almost as low as the frequency of passive constructions. Future constructions are even lower in frequency, as seen in Figure 4 below, using construction frequencies from the COCA.

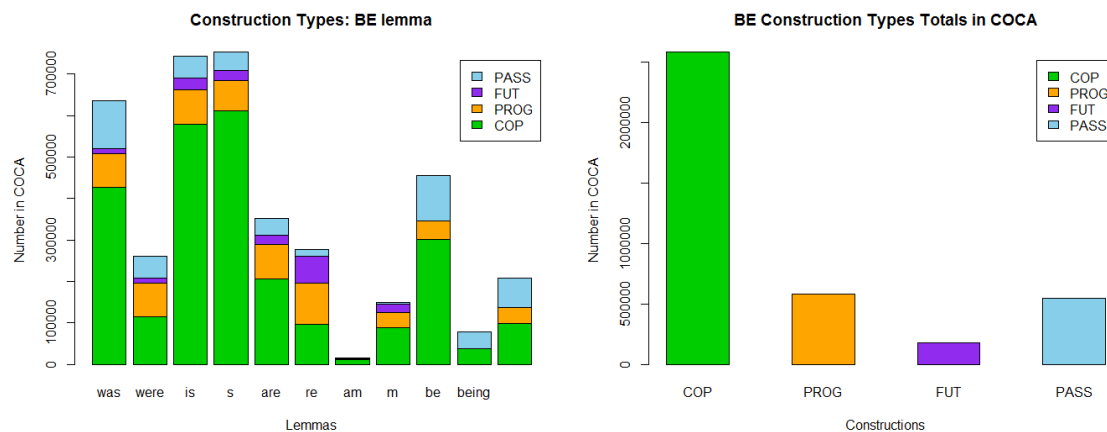


Figure 4. COCA BE Frequency Distributions

Figures 5 and 6 show the inflection and construction frequency distributions in the Buckeye and Redford corpora, respectively. In both corpora, copula constructions far outnumber any other constructions and passives are quite rare. Both have around the same proportion of future constructions, but the Redford Corpus has proportionally more

progressive constructions than either the COCA or the Buckeye Corpus. This is likely due to the genre of the Redford Corpus, which is made entirely of narratives (cf. Chapter III, Section 2), which describe actions in progress in order to move the narrative along. In any case, copulas are far more frequent than any other kind of *BE*, so if construction frequency were to play a role in either word shortening or contraction, we would expect to see the effects strongest with the copula.

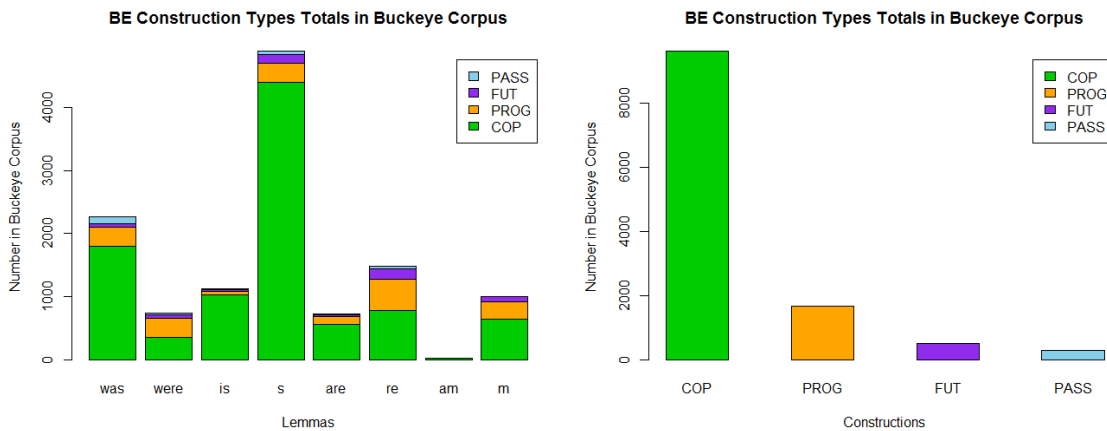


Figure 5. Buckeye Corpus *BE* Frequency Distributions

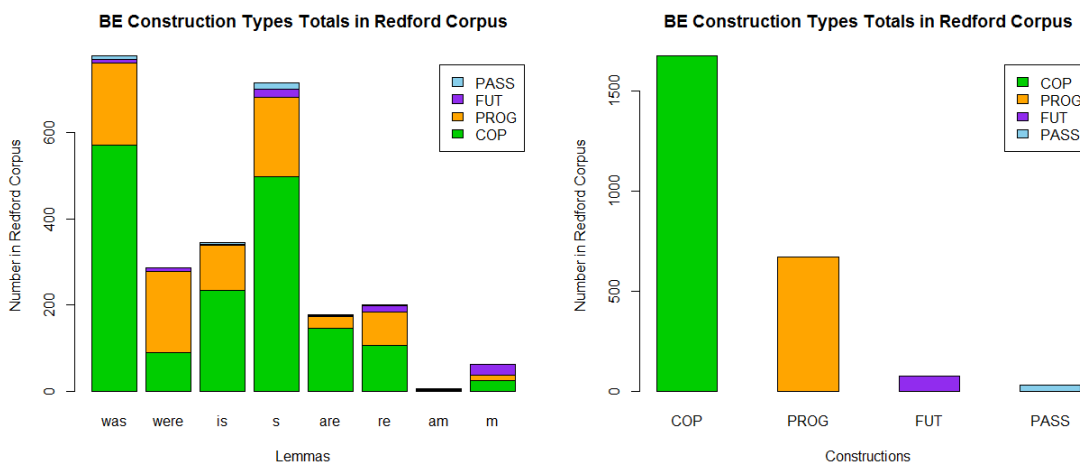


Figure 6. Redford Corpus *BE* Frequency Distributions

2.3. Factors Influencing Reduction Related to Word Class and Constructions

2.3.1. Lexical Class and Frequency

Previous research has shown that grammatical words are more subject to reduction than lexical words (Ansaldo and Lim, 2004; Bell et al., 2009; Swanson et al., 1992; van Bergem, 1995). This could be due to what some authors have argued is a fundamental difference in how lexical and grammatical words are stored and processed by speakers. For instance Ullman (2001, 2004) argues that the mental lexicon is idiosyncratic and dependent on declarative memory but that the mental grammar is regular and so is dependent on procedural memory. He has even argued that these areas are stored in different parts of the brain, with the former located in the temporal lobe and the latter located in the frontal cortex and basal ganglia. It is well known that event-related potentials have shown that difficulties in semantic processing can elicit a N400 peak, which is an increase in electrical activity about 400 milliseconds after encountering the problematic triggering target, and that difficulties in syntactic processing elicit a P600 peak (Ullman, 2001). FMRI studies have shown that lexical and semantic processing is associated with activation in the temporal lobe (Bookheimer et al., 1993; Damasio et al., 1996) and that syntactic processing activates the left basal ganglia (Embick et al. 2000, Stromswold et al. 1996). However Ullman also states that “lexically stored syntactic knowledge” such as knowing the arguments a verb takes (2001) can activate the temporal lobe as well (Kuperberg et al., 2000), which makes the distinction between lexical and grammatical knowledge seem less distinct. Nonetheless, under this kind of “words and rules” approach, frequency effects are thought to influence lexical words but have little to no impact on grammatical words (Ullman 2001). Accordingly, Bybee (2001) argues against this approach by documenting

effects of word frequency on the phonetic reduction of the regular, eminently grammatical, *-ed* past tense suffix.

Research on grammaticalization documents a cline from prototypically lexical to prototypically grammatical words, which is traversed by words as they grammaticalize, with no clear point at which a word ceases to be lexical and becomes grammatical (Bybee, 2001). Instead of talking about lexical and grammatical classes of words, grammaticalization theorists speak of degree of grammaticalization, referring to the synchronic position of a word (in a particular use) on the diachronic lexical-to-grammatical cline. According to this perspective, the processing differences between grammatical and lexical words may be due to uncontrolled and uncontrollable differences between function and content words, the extremely high frequency of grammatical words in comparison to lexical words (Bates and Goodman, 1997; Bell et al., 2009; Bybee, 2001; Bybee, 2007; King and Kutas, 1998). On this account, we expect probability and frequency to influence pronunciation of both “lexical” and “grammatical” units, as documented by Bybee (2001) for *-ed*.

Pierrehumbert (2001) argues that phonological grammar includes not only “the general principles of phonological structure”, but also “word-specific phonetic detail.” This word specific detail is needed in part because of frequency effects on pronunciation. The same morpheme appearing in words with different frequencies can have different realizations. This is because “words which are highly expectable are produced faster and less clearly than words which are rare or surprising” (Pierrehumbert, 2001). In her 2001 application of exemplar theory to phonetics, this means that listeners build up memories of hypo-articulated forms of frequent words, and then in turn use these memories to

produce their own speech. Words that are often lenited will become more lenited in their long-term representations.

There are certainly duration differences in homophonic word forms based on their frequency levels in corpus studies (Bell et al., 2003; Jurafsky et al., 2002; Gahl, 2008; Schuchardt, 1885). For example, Gahl (2008) showed that homonyms such as the more frequent *time* and less frequent *thyme* have different lengths, where more frequent words are shorter. However, she excluded grammatical words from her study. Jurafsky et al. (2002) examined the homophonous (or at least polysemous) grammatical words *to*, *that*, *of* and *you*. Frequency was found to be an explanatory factor for reduced vowel production for *that* and *of*. However, these corpus studies did not specifically compare grammatical items to lexical ones.

An experimental study from van Bergem (1995) did compare homophonous grammatical morphemes to lexical morphemes, in Dutch. The examples used were equivalent to comparing the modal verb *can* to the first syllable of *candy*. He found that grammatical morphemes were shorter and pronounced less distinctly than lexical morphemes. He also found that lexical morphemes were more subject to differences in stress than grammatical morphemes. Swanson et al. (1992) also found that lexical words are more likely to be lengthened when they are spoken to children than adults, whereas the durations of grammatical words are unlikely to change no matter the audience. Other authors have argued that content words typically have longer vowels than function words because they are more likely to be stressed, focused or carefully pronounced (i.e., accented) due to higher information content and higher semantic weight (Aylett and Turk, 2004; Bolinger, 1972; 1975; 1985).

The present study teases apart reduction due to degree of grammaticalization and the reduction due to usage-frequency because for *HAVE* and *BE* because for *HAVE*, the content word *HAVE* is so much more frequent than the function words *HAVE* and for *BE*, while all meanings are function word meanings, the oldest/source meaning is so much more frequent than the newer/grammaticalized meanings.

2.3.2. Construction

In studies of contraction, authors found that construction type (or ‘syntactic context’) was predictive of rates of contraction. Labov (1969) compared *BE* before *gonna*, progressive verbs, noun phrases, adjectives and locatives in AAE. The most contraction and deletion occurred before *gonna* and other progressive verbs. Less contraction occurred before locatives and adjectives and the least reduction was found before noun phrases. Overall, Labov’s data suggest that progressive and future constructions favor reduction over copula constructions, with the caveat that it is not clear which (if any) of the differences reported in Labov (1969) are statistically significant, as no inferential statistics are reported. Four other studies examined the influence of syntactic context on contraction, all in SAE. McElhinny (1993) argues that contraction is most favored before progressive verbs, then before adjectives, then locatives, then before *gonna*, with the least reduction before noun phrases. MacKenzie (2012) distinguishes between verbal, nominal and adjectival following contexts for *BE*, finding more contraction in pre-verbal contexts (progressive and future constructions) than when *BE* is a copula. This finding is also reported by Bresnan and Spencer (2013), who use the same coding as MacKenzie (2012). Thus, overall the literature appears to be in agreement that there is more contraction in progressive constructions than in copula constructions but

there is disagreement about where the future construction falls. Barth and Kapatsinski (In Press) found that contraction is more likely for future constructions and less likely for passive constructions (not included in any of the previous studies), but that rates of contraction of copula and progressive constructions were relatively equal. The present study focuses not only on contraction for these constructions, but on word shortening as well.

2.3.3. Probability

Perhaps more important than frequency for reduction in speech generally is probability.

The probabilistic reduction hypothesis (Bell et al., 2003; Gregory et al., 1999; Jurafsky et al., 2001) states that word forms are more reduced when they are probable given the current context (i.e., predictable). Context includes local context (neighboring words), syntactic or lexical structure, semantic or style expectations, and discourse factors (Bell et al., 2003). In the current study, reduction is operationalized as words being short (in milliseconds) and/or contracted, although other studies of probabilistic reduction have also included consonant lenition, flapping, stop release, vowel centralization, increased coarticulation, and lack of intonational prominence as ways to measure reduction (Aylett & Turk, 2006; Bell et al., 2009; Byrd, 1994; Bybee, 2001; Fowler & Housum, 1987; Gahl, 2008; Gahl and Garnsey, 2004; Jurafsky, 2003; van Bergem, 1995).

Most previous studies of probabilistic reduction have investigated content words, excluding function words due to the assumption of content and function words being processed differently (a la Levelt, Roelofs and Meyer, 1999; Ullman, 2001). Some studies have also found evidence supporting processing differences between content and

function words. For example, function words are less affected by repeated mention (Bell et al., 2009) and more affected by predictability given preceding context (Jurafsky et al., 2001). Nonetheless, there are also strong similarities. For example, both content and function words alike the following context has a strong effect on word duration (Bell et al., 2009). Furthermore, the processing differences may be explainable by other differences between content and function words rather than the function/content distinction. In particular, repetition has weaker effects in frequent words compared to rare words (Forster and Davis, 1984), and preceding context may be more variable for function words (which tend to follow content words) than for content words (which tend to follow function words). Furthermore, in English, some function words, including auxiliaries, cliticize to preceding words, making it unsurprising that they are strongly affected by the units to which they cliticize.

Several studies have examined reduction of function words, and particularly *HAVE* and *BE*. In particular, contraction is more likely when *BE* or *HAVE* occurs in a high-frequency word sequence (Krug, 1998; Bresnan and Spencer, 2013) or is highly probable given the context (Frank and Jaeger, 2008; Bresnan and Spencer, 2013). Krug (1998) finds that singular subjects are more likely to trigger contraction of *BE* or *HAVE* than plural subjects and that first person subjects are more likely to trigger contraction than second person subjects, which are more likely to trigger contraction than third person subjects. He argues that the source of this hierarchy lies in differences in joint probabilities of particular host words and the contracted element. He finds that high joint probability of the target and the word preceding it (its host) predicts contraction better than transitional probability. Barth and Kapatsinski (in press) also found this to be the

case for low frequency words preceding *is*, *am* and *are*, but found that for all preceding words, a simple a difference between pronouns vs. full noun phrases predicts contraction better than a continuous probability variable. Additionally, they found that high transitional probability (and not joint probability) of the target and the word following predicts contraction.

It is clear that there are probabilistic effects in human speech for both content and function words. There are two main types of theories accounting for why these effects exist: speaker-internal accounts and listener-oriented accounts. Proponents of speaker-internal (production-based) models argue that words are reduced when they are highly frequent or highly predictable from context because it takes less time for a speaker to retrieve the word from their own mental lexicon in the course of speech production (Bell et al., 2009; Ferreira, 2008; Gahl 2008; Gahl, Yao, and Johnson, 2012; Seyfarth, 2014) or because of speed of word choice or syntactic construction during syntactic production (Ferreira, 2008; Ferreira and Dell, 2000). Although listener-oriented and speaker-internal models both predict that the words that are quick for a speaker to produce (leading to reduction) are the same words that are easy for a listener to understand (also leading to reduction), studies of the effects of high density phonological neighborhoods (Gahl, et al., 2012; Munson, 2007) provide evidence for a speaker-internal account. When a word has many phonological neighbors, associated activations between semantic and phonological forms spread to neighbors. This makes words with dense neighborhoods more difficult to perceive because of competition from many neighbors. However, high neighborhood density makes lexical access in production easier, as assessed by reaction time in picture naming (Vitevitch, 2002). Word duration appears to pattern with production difficulty,

reflecting the speaker's, and not the listener's ease of lexical access. Thus, words are shorter in dense neighborhoods despite being more difficult to understand (Gahl, et al., 2012; Munson, 2007). Since words with more neighbors present more competition for a listener to decode which word they are hearing, it would be helpful to a listener if words with more neighbors had longer, clearer signals, which, evidently, they do not always have (Gahl et al., 2012)². But note that Jaeger and Tily (2010: 326-327) stress that the time to produce a word and the time to plan a word are not necessarily linked under a probabilistic view of speech production:

While longer latencies would be expected if less predictable word tokens are harder to produce, longer *durations* would only be expected if each segment (phone) in less predictable word tokens is harder to retrieve than the segments in more predictable word tokens. In other words, longer durations for less predictable tokens are only expected if the segments of less predictable word tokens are on average less predictable than the segments of more predictable word tokens (emphasis in original).

Proponents of listener-oriented (intelligibility-based) models argue that words reduce when the speaker assesses them to be easy for a listener to predict from context. Less signal (in terms of milliseconds of duration) is necessary for a listener to decode the meaning of the word and therefore understand the speaker's intent, so a speaker can get away with a shortened, reduced pronunciation, saving effort. Words are longer when a word may be difficult for a listener to understand because it is infrequent, difficult to predict from context or highly confusable with similar other words (i.e., has a dense

² Several studies using experimental data or different measures of neighborhood density have shown the opposite, that words with high neighborhood densities are produced more clearly (Munson, 2007; Munson and Soloman, 2004; Watson and Munson, 2007; Wright, 2004, inter alia)

phonological neighborhood). Speakers adjust their speech in speed, coarticulation, consonant lenition, vowel centralization and vowel choice, among other adjustments, based on what they perceive as listener-needs (Arnold, 2008; Aylett and Turk, 2004; Clark et al., 1991; Clark et al., 1987; Flemming, 2010; Fox Tree and Clark, 1997; Freed, 1978; Ferguson, 1977; Galati and Brennan, 2010; Levy and Jaeger, 2007; Lindblom, 1990; Lockridge and Brennan, 2002; Pluymaekers, Ernestus and Baayen, 2005; Swanson et al., 1992; van Son and Pols, 2003). Function words, in particular, are high in frequency and lack much semantic content, meaning they can be reduced or deleted because they are not essential for a listener to process the information in a speaker's utterance, or are recoverable from context (Krug, 1998). Bates and colleagues (Bates et al., 1991; Bates, Devescovi and Wulfeck, 2001; Chen and Bates, 1998) have argued that Broca's aphasics omit function words for this reason (also cf. Pick [1913]). Because production is oriented to the listener in a particular context in each speech act, words extracted from conversation and presented to listeners in isolation tend to be difficult for listeners to decode (Ernestus et al., 2002; Tucker, 2011; Van de Ven et al., 2011; Pickett and Pollack, 1963). Due to word reduction, context is needed for the words to be successfully recognized or decoded (Ernestus and Warner, 2011; Shockey, 1998; Warner et al., 2009). Researchers have found that listener accommodation takes place at many levels of language including discourse variation, such as the inclusion or omission of a narrative element, number of words and detail (Galati and Brennan, 2010; Lockridge and Brennan, 2002), not only in word production.

An information-theoretic account of probabilistic reduction is presented by Uniform Information Density (UID) theory. Words that are more informative are longer

than words that are less informative to maintain a “uniform information density” so that over a period of time roughly the same amount of information is transmitted. (Fenk and Fenk, 1980; Frank and Jaeger, 2008; Jaeger 2010; Jaeger and Tily, 2010; Zipf, 1935). As an example, in a speech segment of 500 milliseconds, a speaker may produce one “informative” (unpredictable) word, such as *bees*, or two uninformative (or predictable) words such as *that one*. Essentially, the idea behind UID is that speakers want to transmit information to their listeners and prefer to spend time on informative segments of their utterances rather than uninformative ones, because they want to maintain a fairly consistent transfer of information over time, rather than lengthening and shortening to highlight the importance of informative words and signal the unimportance of predictable words. Researchers have found that uniform information transmission is preferred at levels of language beyond pronunciation variation, including syntactic variation such as the inclusion or omission of relativizers in relative clauses (Levy and Jaeger, 2007; Wasow et al., 2011).

Unlike more ‘active’ listener-oriented models such as H&H Theory (Lindblom, 1990), UID does not commit to the speaker constantly monitoring the listener’s state of mind. Uniform Information Density may also be an outcome of adaptation and long-term accommodation over the course of multiple interactions with a listener rather than online adjustment of signal clarity based on moment-to-moment estimation of common ground. Minimization of online computation appears to be a recent trend in listener-oriented theorizing in response to concerns about memory and processing time demands of listener modeling. Thus, working in an audience design framework, Galati and Brennan

(2010) propose a one-bit model of the listener where the speaker only keeps track of whether the listener has heard the narrative before at a relatively coarse level.

Some studies have found that although lexical reduction is influenced by accommodation of a listener, articulatory reduction is not (Bard et al., 2000; Bard and Aylett, 2004), but other studies have shown that both speaker experience and estimation of addressee knowledge or task importance affects word shortening (Gregory et al., 2001), as well as prominence measured by pitch (Watson, Arnold & Tanenhaus 2008).

In the present study, either a speaker-internal or listener-oriented account would predict the same basic behavior for the words under investigation: namely that words will be more reduced (shortened or contracted) when they are probable and words will be less reduced (lengthened or uncontracted) when they are less probable. As it is a corpus study, the present work does not provide information on whether the speaker engages in online modeling of the listener or whether this modeling affects articulation. However, if any kind of listener accommodation constitutes support a listener-oriented framework, without evidence for online listener modeling being necessary, then my results on speech directed by the same caregiver speakers to children of different ages fit the bill. The same speakers appear to speak differently to the same interlocutor as the interlocutor ages, in a way consistent with the effects of age on child language knowledge.

2.3.4. Reduction Begets Reduction

If factors often facilitate efficient production or high intelligibility for particular words in particular contexts, those words can retain the effects of those factors even outside of those contexts. For example, Baese-Berk and Goldrick (2009) found longer

VOT of initial voiceless consonants in words that had a minimal pair with an initial voiced consonant (pox vs. box) but not when a word lacked a corresponding minimal pair (posh vs. *bosh). Importantly, this effect persisted even when the minimal pair did not occur in the stimulus set presented to participants (Peramunage et al., 2010). Dilts (2013) found that adverbs that shorten more when they occur after words with parts-of-speech that typically precede adverbs. Torres Cacoullos (1999) found that Spanish auxiliary-gerund sequences diachronically increased in frequency and reductive change, but that the construction frequency also interacted with register to influence reduction. In Spanish, the kinds of verbs that occur in constructions with less reduction (stand alone gerunds encouraging postposed clitic position as opposed to *-ndo* constructions) are associated with formal contexts, so much so that over time the postposed clitic has itself become associated with formality. Raymond and Brown (2012) find influence on reduction of word-initial /s/ in Spanish from both online production factors and factors due to cumulative exposure to reduction-likely contexts. Seyfarth (2014) found that content words that often occur in highly probable context tend to be reduced, even when in an unpredictable context. These kind of findings indicate that speakers are sensitive to tendencies associated with words because they produce more intelligible speech (i.e., voicing contrast) or more efficient speech (i.e., word shortening) even when the current context does not require it. Bybee and Torres Cacoullos (2008) argue this can be accounted for by an exemplar account, where words that tend to occur in reduced contexts have a greater proportion of reduced exemplars in their exemplar cluster than words that do not tend to occur in reduced contexts, making a reduced exemplar more likely to be selected for production.

In the present study, we are interested in a particular kind of context: construction context. If the construction that *BE* or *HAVE* occurs in usually makes *BE* and *HAVE* predictable and therefore reduced, does that reduction persist in the particular construction even when the particular instance of *HAVE* or *BE* is less predictable? If so, then speakers are keeping track of and storing phonetic information as a characteristic of the construction, and if this is so, we then can ask if children do it or if it is only adults, who have had enough experience to learn the phonetic peculiarities of constructions and may be better at tracking probability in context. If not, then findings would support online calculation of probability in each instance and would reject abstraction over construction productions.

2.4. Factors Influencing Reduction Due to Speaker Characteristics

2.4.1. Speech Rate

Speech rate, or tempo, is an important factor to consider when examining reduction because in adult speech, faster speech rate is associated with short vowels, particularly shortened stressed vowels (Gay, 1978, 1981; Redford 2014) meaning that reduction will be at least partially dependent on tempo. In child speech, faster speech rate is associated with both shortened vowel and consonant durations (Redford, 2014), not just shortened vowels.

Older adults' speech rates are just slightly slower than younger adults, and gender differences in speech rate are below the just noticeable difference threshold (Quené, 2007) after phrase length is taken into account (but cf. Bell et al., 2003 and Yuan et al., 2006). Speech rate increases as utterance length increases (Quené, 2008) and both

children and adults pause for a longer time before longer utterances (Redford, 2013). Speech rate also varies by region (cf. Kendall, 2013 and Jacewicz et al., 2009 for the United States, Quene, 2008 for the Netherlands). Words are lengthened phrase-finally as well (Oller, 2008; White, 2002; Yuan et al., 2006), and the magnitude of phrase-final lengthening also varies across speakers (Snow, 1994).

Using the Buckeye Corpus (Pitt et al., 2007), Dilts (2013) found that when speakers spoke quickly, they shortened words and deleted segments to a higher degree. Speakers with high average speaking rates showed more segment deletion than speakers with slower average speaking rates, but the reverse was true for word duration. He attributes this to faster speakers having more even speaking rates across a conversation and slower speakers shortening their words to a greater degree when they *do* speak quickly.

Speakers also vary their speech rate to accommodate their interlocutors based on interlocutor characteristics (Kendall, 2013) and whether the content of the utterance is unpredictable or important (Nooteboom & Eefting, 1994; Zwaardemaker & Eijkman, 1928).

Child age obviously has a huge impact in speech rate. As children get older, they are able to increase their speaking rates substantially until they are about 13 or 14 years old, with the greatest increases seen from 5 to 8 years old, due to better motor control (Kowal, O'Connell, and Sabin 1975; Redford, Forthcoming) and children's segment durations are longer when they are younger (Smith, 1978; Kent and Forner, 1980; Lee et al., 1999). As children get older, because of their increase in speech rate, they are able to say more in the same amount of time (Redford, 2013; 2014). Along with an increase in

the fine motor control necessary for speech, during this same time the complexity of syntactic structures increases in child speech (Brown, 1973; Tomasello, 2003; Tomasello, 2008) alongside an increase in working memory (Baddeley, 1986; Gathercole & Baddeley, 1993).

2.4.2. Age

In adult speech, Gahl et al. (2012) did not find a significant difference in duration reduction or vowel dispersion in the speech of younger vs. older adult speakers using the Buckeye Corpus (Pitt et al., 2007), although faster speech rate was associated with shorter durations. We know little about the differences in probabilistic reduction based on age in child speech. This dissertation seeks to fill that gap. Child speech and child-directed speech are particularly interesting for examining function words because although function words are highly frequent in infant-directed and child-directed input, they are acquired late in comparison to content words.

The three construction types for *HAVE* are ‘possessive’, in which *HAVE* is lexical, ‘modal’ and ‘perfect’, in which *HAVE* is grammatical. Some authors have argued that modals are semi-grammaticized forms (Bybee 1985; Plank 1985; Palmer 1986) meaning they are something between a lexical and grammatical word. For the present study, I consider the semi-modal auxiliary to be a grammatical word as it takes a dependent verb. Example sentences from the children’s narrations used in the present study are:

14) possessive: *They had a pet frog named Bob.*

15) modal: *It sank the boat and his mother had to get it for him.*

16) perfect: *The boy was real sad to find that the frog had left.*

The lexical, possessive verb *HAVE* is produced in appropriate contexts by age 3. Some examples from Diessel (2004:102) show its use by a child at ages 2;9 and 2;10: *He sayed he has something to play with for me* and *She gonna say I have a pretty dress on*. Tomasello (1998:354-355) reports the use of possessive *have* by one child at age 1;7-10: *Have-it, Girl have that umbrella, Have juice in my bottle* and *Have a donut for you*.

The perfect auxiliary *HAVE* is generally used appropriately around age 5 according to Brown (1973), but Diessel (2004:64 and 100) finds perfect constructions in the speech of a 3;8 and 4;6 year old: *Children have begin to sing* and *I see you have bought new toys* respectively. The modal verb is generally used correctly around age 2;3 (Diessel, 2004:74), but initially as a pivot (Braine, 1963) in the form of *hafta*. The related *gotta* is acquired generally around age 2;6 (Diessel, 2004:74). There are variable specific ages of acquisition for these constructions, but they can be ranked in a given child's acquisition as follows: 1) lexical possession, 2) semi-modal auxiliary and 3) perfect auxiliary.

Children acquire present tense forms of *BE* by age 3 and past tense forms by age 5. Copulas and progressive auxiliaries are acquired by Brown's (1973) stage II (2-3 year olds). Before that, children simply omit copulas, auxiliaries and the *-ing* of the participle of progressive constructions. In progressive constructions, children start producing the full participle, with *-ing*, before they include the auxiliary (Brown, 1973:253). Passives come later, but are acquired before age four (Brown, 1973). Brown's data (1973:259) also shows that children produce non-contracted copulas and auxiliaries before they produce contracted ones, perhaps because parents in his study did not contract copulas or auxiliaries very often (359).

While *BE* and *HAVE* are used in appropriate contexts early, it takes longer to acquire adult-like patterns of contraction, and in dialects that permit it, omission. Kovac and Anderson (1981) examine the deletion and contraction of *BE* in the speech of children, from groups matched for age (3 year olds, 5 year olds, 7 year olds), race (black, white) and income-level (middle-income, lower-income). *BE* deletion is a well-known feature of African American English and is more likely after vowels, after pronouns and when it is a progressive or future auxiliary, as opposed to a copula (Labov 1969). The youngest children in Kovac and Anderson's study were still acquiring copula and auxiliary usage, sometimes omitting *BE* for developmental reasons, rather than to match adult non-standard dialects. However, the oldest children in the study reflected adult usage patterns, although not fully. There was a decrease in deletion and an increase in contracted (*'s*, *'m*, *'re*) and non-contracted (*is*, *am*, *are*) forms over the age cohorts of white middle class children, which is appropriate to Standard American English. This pattern was the same for most white working class children as well. Black working class children learn deletion earlier than black middle class children, probably due to higher rates of *BE* deletion from their parents, and then later from their peer groups. Two white working class children in the 7 year old cohort also had high rates of *BE* deletion, probably due to influence from their peers. Kovac and Anderson (1981) also examine the contexts for *BE* deletion and contraction for the black children, finding that their contraction patterns match adult usage, but that their deletion patterns do not. However, for some of their linguistic variables there are fairly low *n*'s, so with more data, it may be that child usage patterns would be more similar to adult usage patterns.

In the child speech portion of the data, it is necessary to be aware that until at least ages 7-12, children have a great deal more within-category acoustic variation in their speech than adults (Foulkes and Docherty, 2005; Lee, Potamianos and Narayanan, 1999; Smith and Kennedy, 1994; Smith, Kennedy and Hussain, 1996, Snow, 1995) and some aspects of speech motoric development persist past adolescence (Cheng et al., 2007).

Children's durations of reduced (unstressed) syllables, while relatively shorter than prominent (stressed) syllables, are much longer than reduced syllables in adult production (Allen and Hawkins, 1980; Kehoe, Stoel-Gamon, and Buder, 1995) and the variability in unstressed syllable duration is higher than in stressed syllable duration (Goffman, 1999). Particularly, producing the very short vowel durations associated with unstressed syllables is difficult for children, even at age seven (Allen and Hawkins, 1980; Ballard et al., 2012). Although they can successfully produce relatively shorter vowel durations in unstressed syllables in relation to stressed syllables, their short vowel durations are still much longer than those of adults (ibid). This may indicate that school-aged children will still be producing very long function words, which are often unstressed monosyllables, particularly when they occur before a stressed syllable.

Even when children's production of stressed syllables is perceived as correct, the kinematic movements underlying those productions are slower for 3;10-4:9 year olds than for adults (Goffmann and Malin, 1999). Motor control is still developing in late childhood, until 12-16 years of age (Lee, Potamianos and Narayanan, 1999; Smith, 2006). Until then, children have a significant amount of variability and differences in magnitude of their segment durations until age 12 (Lee et al., 1999) and even after that

there is further development in their consistency in oral-motor coordination (Smith, 2006).

Children do not reduce function words as much as adults do (Allen and Hawkins, 1978; Goffman, 2004; Redford, Forthcoming; Sirsa and Redford, 2011) and children that fail to reduce function words in relation to content words while reading aloud are perceived as reading less fluently (Lord, Berdan, and Fender, 2009). Second language speakers who reduce function words to a lesser extent are also rated as sounding less proficient (Baker, Baese-Berk, Bonnasse-Gahot, Kim, Van Engen, and Bradlow, 2011). As children get older, they find that in their own speech production, only certain parts of a message can be reduced while maintaining intelligibility. The present study seeks to address whether there is a wholesale reduction in all instances of function words over increased age (due simply to their undisputed better production of unstressed syllables) or whether the reduction is influenced by meaning, context and probability, indicating an additional understanding of what is important and unimportant in an utterance.

2.4.3. Gender

Studies of probabilistic reduction in adult speech have conflicting results as to whether there is a significant difference between the genders. Using a corpus of read speech, Byrd (1994) found that men spoke faster than women, used more taps and syllabic [n], but did not centralize vowels to a greater degree than women. Bell et al. (2003) also found that men spoke faster than women using a corpus of spontaneous speech and that there was a significant interaction between age and gender that older women spoke more slowly than old men. Gahl et al. (2012) did not find a significant difference in duration reduction or vowel dispersion in the speech of males vs. females.

Using data from the Buckeye corpus, Dilts (2013) found that although word duration was unaffected by gender, men were more likely to delete segments than women.

Additionally, he found that the gender of the interlocutor (interviewer) had an effect on segment deletion. Over the course of an interview, speakers tended to increase their rate of segment deletion with male interviewers and decrease segment deletion with female interviewers, particularly for frequent words, showing sensitivity to characteristics of their listener.

Studies of child-directed speech have shown strong differences between male and female caregivers. When comparing speech to adults v. preverbal infants in French, Italian, German, Japanese, British English and American English, Fernald et al. (1989) found that all parents used higher F0 means, minima, maxima, and variability as well as shorter utterances and longer pauses when speaking to infants. Only mothers used a wider F0 range. Of all the language groups, American English speaking mothers had the greatest intonational exaggeration in their infant-directed speech. For F0 range, F0 hertz were converted into semitones, which is a logarithmic transformation of the difference in hertz values. Rondal (1980) found that like mothers, fathers simplified their speech when speaking to toddlers (1;6-3;0) in French, shortening their MLUs even more than mothers, matching average MLU of the children they were speaking with. In addition to differences between adult males and females in caregiver speech, some caregivers speak differently to their child, depending on whether or not the child is male or female (Foulkes, Docherty and Watt, 2005). Additionally, there may be some interaction between parent and child gender. Reese, Haden and Fivush (1996) found that mothers made distinctions in their speech depending upon whether or not they were speaking to

boys or girls, but that fathers did not. Caregivers' use of features of infant directed speech taper off by age five (Bellinger, 1980) and the child's need for this speech from caregivers diminishes over time (Garnica 1977) and so gender is unlikely to be a factor influencing function word reduction in caregiver speech.

2.5. Additional Factors Influencing Reduction

2.5.1. Preceding Phonological Context

Studies of contraction have found that the host word for a contraction has an effect on its likelihood of contraction. Host words that end in a vowel, as opposed to consonants, are favorable for contraction (Labov, 1969; MacKenzie, 2012; McElhinny, 1993) or deletion in African American English (AAE) (Labov, 1969). These same studies found that pronoun host words were favorable to contraction, which should be expected as many pronouns end in vowels. Barth and Kapatsisnski (In Press) found that pronouns were predictive of contraction, but that neither the preceding nor following phonological context had a significant effect on rates of contraction once word type was taken into account.

2.5.2. Priming

Studies have shown that a speaker's use of a particular variant can prime the further use of that variant over another. Poplack (1980) found that plural –s deletion in Spanish is more likely if the preceding instance of plural –s within the same phrase is also reduced. Scherre and Naro (1991) found that agreement marking or non-marking was likely to prime further agreement marking or non-marking across different phrases within the same clause in Brazilian Portuguese. Cameron and Flores-Ferrán (2004) found that

speakers of regional dialects of Spanish preserve the same kind of subject expression (null or pronominal) over clauses, as did Travis (2007) in other Spanish varieties. Torres Cacoullos and Travis (2013) find that American English speakers are much more likely to use the rare strategy of an unexpressed first person subject when it is preceded by another unexpressed first person subject. Barth and Kapatsinski (In Press) found that American English speakers are much more likely to contract *is*, *am* or *are* after contracting one of these words earlier in their utterance, or to produce non-contracted forms if they had recently produced non-contracted forms of these words.

2.5.3. Disfluencies

Speakers lengthen word durations before disfluencies, presumably in order to buy time to access the word they trying to find (Bell et al., 2003; Schachter et al., 1991; Kapatsinski, 2005; Schnadt, 2009). Disfluencies also often occur before low probability words more than before high probability words (Beattie & Butterworth, 1979; Goldman-Eisler, 1957; Maclay & Osgood, 1959; Schachter et al., 1994; Tannenbaum et al., 1965), resulting in longer durations (Tily et al., 2009), showing that high probability positively affects ease of word access.

2.6. Conclusion and Roadmap

The present study examines in detail whether there are effects of probabilistic reduction on the production of the function words *BE* and *HAVE* in adult speech, child-directed speech and child speech. I find that there are effects of probabilistic reduction in the speech of all of these groups, and these effects strengthen in child speech as children

get older. Strengthening of sensitivity to probabilistic factors indicates that children, in addition to progressing in speech ability, also progress in language production during their school-age years. Children become more aware of what is (un)important in their speech and then use this information to be more efficient speakers.

In Chapter III, I first examine in detail reduction of *BE* and *HAVE* in the Redford Corpus, examining the speech of children and their caregivers. I show children are sensitive to probability in word shortening and contraction. In this chapter I also show that speakers have styles of reduction, that is, a strategy for the compression of unimportant information, and that children develop a style around age 7.

In Chapter IV, I confirm the results of the probabilistic reduction of *BE* and *HAVE* from Chapter III in a larger corpus: the Buckeye Corpus. In this chapter I also examine the probabilistic effects of construction types and show that even though copula and possessive constructions are more frequent than other construction types, they also have lower average following transitional probability. I show that constructions that reduce to the highest degree have higher average following transitional probability, showing that speakers are more sensitive to *probability* as a feature of constructions than they are to *frequency* as a feature of constructions. Crucially, I show that effects of high average probability increase reduction, even in specific contexts of low probability, meaning that an association between high probability and reduction is at least partially stored in the mental lexicon, not calculated online during language production.

Finally in Chapter V, I bring the results of the above studies together to show how probabilistic reduction affects language even at low levels, like function word production, and that sensitivity to probabilistic factors is a sign of language proficiency. I also argue

that average word predictability captures the intuition of grammaticalization researchers that phonetic erosion accompanies grammaticalization due to the restricted contexts that a grammaticalized word can occur in.

CHAPTER III

REDFORD CORPUS: REDUCTION IN CHILD AND CHILD-DIRECTED SPEECH

3.1. Introduction

The present chapter examines probabilistic reduction in child speech (CS) and child-directed speech (CDS) using data from the Redford Corpus. Children's speech is quite different than adults' speech. It is more variable, varies in less pattern-based ways, and has a more restricted inventory of lexical items and grammatical constructions, among other differences (Foulkes and Docherty, 2005). Much of the speech children hear is not like inter-adult speech. Although produced by adults, it is directed towards children and has its own special characteristics: simplified syntax and vocabulary, repetitions, slower speech rate, longer pause durations, larger pitch range, additional emphasis on lexical items, but no additional emphasis on grammatical items (Fernald et al., 1989; Gallaway and Richards, 1994; Snow, 1995; Swanson et al., 1992). Yet, all children acquire adult-like speech and adult-like patterns of variation. In order to examine the progression from child speech to more adult-like speech, I examine here patterns of probabilistic reduction from children ages 5-10 and compare them to patterns of probabilistic reduction in CDS.

I first discuss the Redford Corpus in some detail in section 3.2 and the methodology of the study in 3.3 including dependent and independent variables and the statistical analyses used: random forest variable importance ranking and multimodel

inferencing for significance testing. In sections 3.4-3.7 I present the results of the studies of probabilistic reduction on word shortening and contraction in CS and CDS. In section 3.8, I present a discussion of the results showing that there is indeed development in probabilistic reduction in CS from ages 5-10, and that caregivers adapt their speech to children in this age range as well.

3.2. Data

Child Speech and Child-Directed Speech data come from data collected by Professor Melissa A. Redford as part of a large-scale, longitudinal study examining the development of prosody of school-aged children³. As one part of a larger study involving a battery of tests⁴, children narrated a picture book called *One Frog Too Many* (Mayer and Mayer, 1975), *A Boy, a Dog and a Frog* (Mayer, 1967) or *Frog, Where Are You?* (Mayer, 1969). The books are approximately twenty pages long but contain no text, so children spontaneously produced a story to accompany the pictures. The caregiver was present with the child and narrated the book first, followed by the child. The caregiver narrated the book once more, followed by the child once more. Sometimes a research assistant narrated the book in alternation with a child, instead of a caregiver. Narrations produced by research assistants are not included in the present corpus. Generally, the second story was transcribed either by the research assistants in the Redford lab, or by the present author. Some instances of the first story were also transcribed. For the current

³NICHHD grant Award Number R01HD061458

⁴ The tests were not given by the author. Tests were given by Prof. Redford and her research assistants from 2009-2011.

study, a selection of 216 stories are used, resulting in a 77,822 word corpus of picture book narrations from children and their caregivers, hereafter the REDFORD CORPUS.

The children in the study were typically developing, as reported by their caregivers, in school, and all spoke a west coast variety of American English (Redford, 2014). Not all caregivers spoke this dialect and for some caregivers, English was not their native language. The data from the participants in these caregiver-child dyads was checked carefully for outliers, and if outliers occurred, those data points were excluded from analyses.

Narrations in the present corpus are from typically developing, hearing children. The corpus of “frog stories” is made up of 146 narrations from 58 children (34 female, 24 male) and 70 narrations from 35 parents⁵ (28 female, 7 male) at three different time points (in 2009, in 2010 and in 2011), allowing us a longitudinal view of the development of variation. Not all stories were transcribed at the time of data collection for this project; therefore for some children there are only data from two time points, or even only one time point. The range of ages of the children is 62 months-130 months. For the specific time points the ranges are 62-107 months for the 2009 cohort, 63-118 months for the 2010 cohort and 63-130 months for the 2011 cohort. Most stories were produced by children between the ages of 90-110 months, that is (7;5-9;2 years old) as shown in Figure 7. The length of the stories ranged from 1 minute 2 seconds to 8 minutes 3 seconds for children and from 1 minute 47 seconds to 6 minutes 47 seconds for adults.

⁵ More children and caregivers participated in the Redford Lab studies than those that have narrations in the corpus in the present study. As the project focused on child speech, many more child narrations were transcribed than adult narrations.

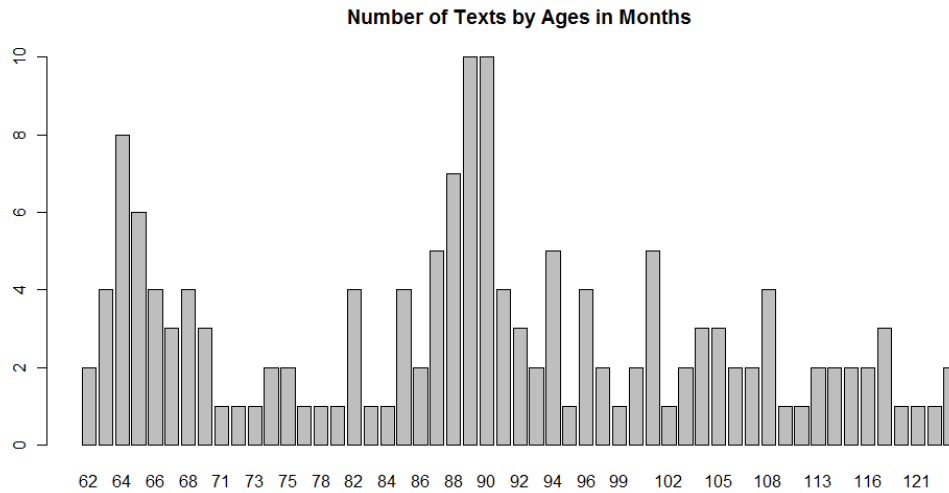


Figure 7. Numbers of stories by age of children in months

3.2.1. Word and Bigram Frequency in a Small Corpus

The Redford Corpus is relatively small and has a different distribution of words and bigrams than a large corpus such as the Corpus of Contemporary American English (Davies, 2008-) or COCA. The COCA contains adults' speech and writing from television news programs, radio programs, newspapers and books and has over 450 million words. Table 1 compares the ten most frequent words and bigrams in COCA vs. in the Redford Corpus. From this table we see that there are some similarities: grammatical words are the most frequent in both corpora, but the Redford Corpus is heavily skewed toward frogs. We also see bigrams that are associative with narrative story-telling, i.e. *and then*, *and he*, *and they*.

Table 1. Ten most frequent words and bigrams in the COCA and in the Redford Corpus

	COCA word	COCA bigram	Redford Corpus word	Redford Corpus bigram
1	the (22,038,615)	of the (2,551,888)	the (7,444)	and the (1,255)
2	be (12,545,825)	in the (1,887,475)	and (5,414)	the frog (1,159)
3	and (10,741,073)	to the (1,041,011)	frog (2,848)	and then (752)
4	of (10,343,885)	on the (861,798)	he (2,817)	the boy (724)
5	a (10,144,200)	and the (676,658)	a (1,929)	and he (681)
6	in (6,996,437)	to be (648,408)	to (1,566)	the dog (655)
7	to (6,332,195)	for the (578,806)	they (1,522)	the little (505)
8	have (4,303,955)	at the (561,171)	was (1,513)	in the (477)
9	I (3,978,265)	in a (498,217)	then (1,214)	and they (446)
10	it (3,872,477)	with the (455,367)	boy (1,194)	frog and (400)

The advantage of using the COCA for calculating word and bigram frequencies and probabilities is that there are over 450 million words in the corpus, meaning the frequency measures are fairly reliable. If something occurs rarely, we can be confident it is indeed rare, rather than this being merely an artifact of a small corpus. The Redford Corpus has many sentences using the word *frog* (and *turtle*, *dog*, etc.), and the COCA has very few, meaning it may potentially be less reliable for predicting the behavior of words like these that are prominent in the Redford Corpus. However, because the Redford Corpus is so small, there are many words and bigrams that occur only once. They are what are called *hapax legomena* and they are unreliable for predicting wider probabilities. With a small corpus, we do not know if these words/bigrams are truly rare,

or if the small corpus just happens to lack them by chance. Therefore, throughout the next sections, I use both frequency information that is task-based, from the Redford Corpus, and frequency information that is more reliable due to size, from the COCA. Dilts (2013) found that local frequencies from the Buckeye Corpus predicted reduction in the Buckeye Corpus better than frequencies from the COCA, so it is possible that local context information is more predictive for its specific context.

3.2.2. Data Structure Visualization in Redford Corpus

We know that children have smaller vocabularies than adults, and we know that children's vocabularies are made up primarily of frequent words. Using the Fruchterman-Reingold layout algorithm in NodeXL (a Microsoft Excel Add-In for the exploration of network graphs), we can get a sense of the structure of the Redford Corpus and differences in the structure of the child v. adult portions of the corpora. The Fruchterman-Reingold algorithm is a force-directed graph drawing algorithm that displays vertices and the edges between them (Fruchterman and Reingold, 1991). Nodes are placed closer together if they have a strong attractive force and placed farther from each other if they have a strong repulsive force. For the present data, the calculation of force by the algorithm is based on how often words co-occur in a string, given all possible co-occurrences (i.e., joint probability). In the figures that follow, each bigram in the Redford Corpus is represented as two vertices connected by an edge. Bigrams that are often used in the corpus (i.e., have a high joint probability) have a strong attractive force and therefore shorter edges. Bigrams that are less often used in the corpus, given all possible combinations, (i. e., have a lower joint probability) have a repulsive force between the

words of the bigram and therefore longer edges between those words. A loop represents a word that has been repeated, because the edge connects a bigram made of two of the same words. The exact placement of nodes depends on the placement of all other nodes and their edges, resulting in the specific shape of the structure. Figure 8 shows the structure of both the child speech (CS) and child-directed speech (CDS) portions of the corpus and Figures 9 and 10 compare the CS and CDS portions of the corpus. There are more child stories than parent stories, so Figure 9 is based on a sample of data from the CS portion of the corpus equal to that of the CDS portion of the corpus used in Figure 10.

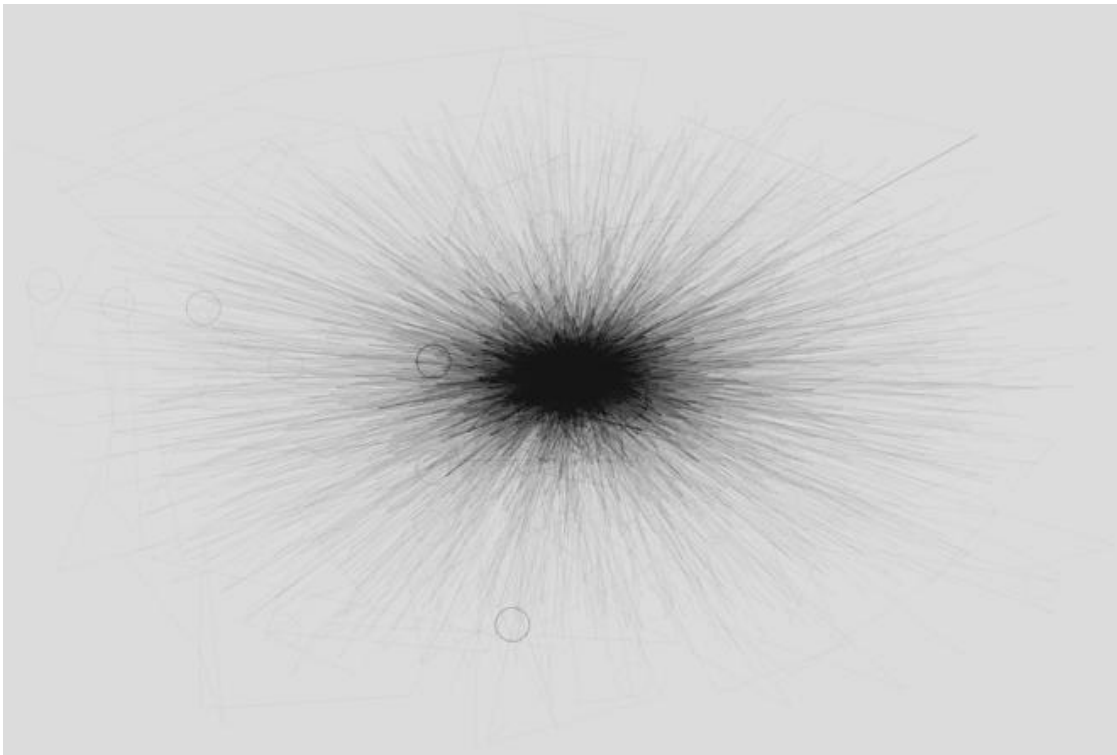


Figure 8. Structure of the Child Speech and Child-Directed Speech Corpus.

Notice in Figure 9 below, where we see the structure of the CS portion of the corpus, that there is a high concentration of highly frequent bigrams. In Figure 10, where we see the CDS portion of the corpus, there is a higher number of infrequent bigrams. In

Figure 9, notice that after a child has used a less frequent bigram, the following word will usually be part of a frequent bigram again. This is represented by edges from the vertices far from the center returning to the center of the figure to be connected to vertices with short edges. In Figure 10, we see that vertices with long edges are more often connected to other vertices with long edges. This represents that an adult will use one infrequent bigram after another. Additionally, the center is far less dense and the dark black center part is relatively smaller than in the one in Figure 9. This represents that the adults use the highly frequent bigrams less often than children.

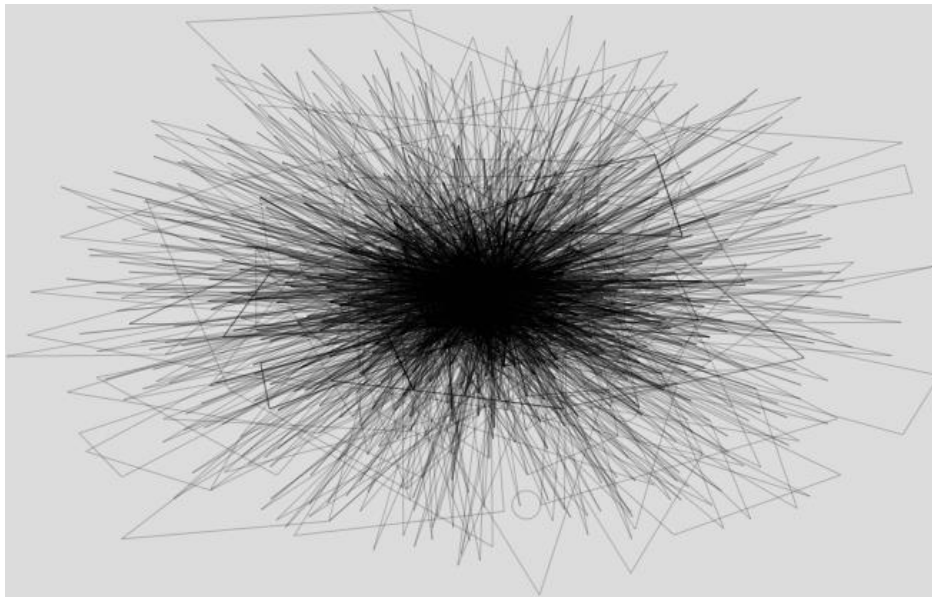


Figure 9. Structure of the Child Speech Texts

In Figures 11 and 12 we see the same text structures as in the previous two figures, but with a red line connecting a selection of the vertices. This red line represents one narration from a child (Figure 11) and one from an adult (Figure 12). Figure 11 allows us to better see that a child will often use highly frequent bigrams, and when they do use a less frequent bigram, will then follow it with another highly frequent bigram, as in a pivot grammar for a young child (Braine, 1963; Tomasello, 2003). Figure 12 shows

us that an adult uses more bigrams with lower frequency and may use a few less frequent bigrams, one after another.

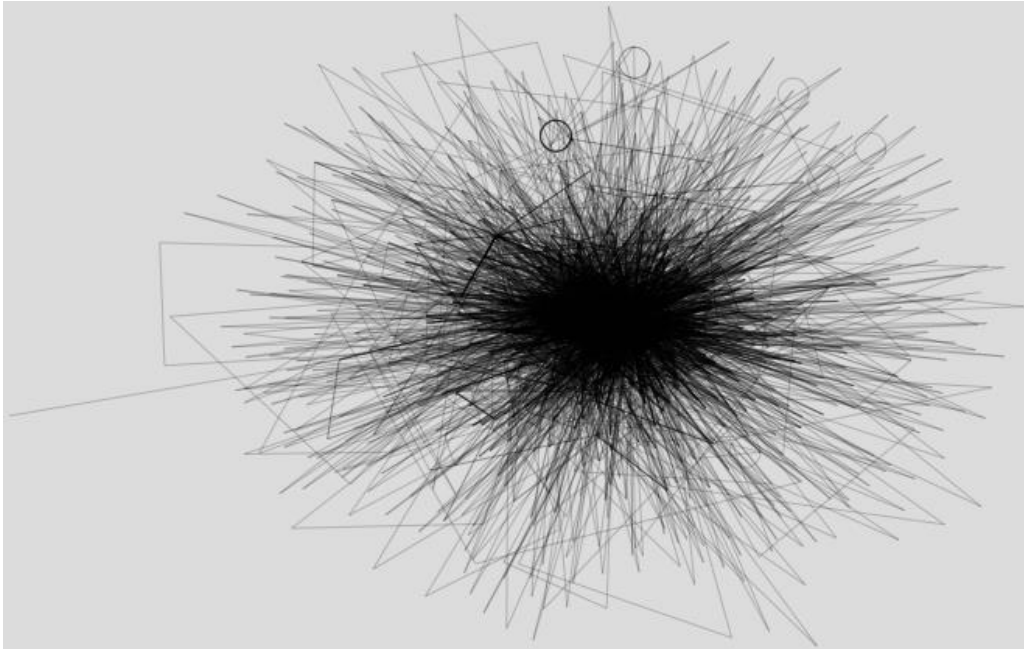


Figure 10. Structure of the Child-Directed Speech Texts

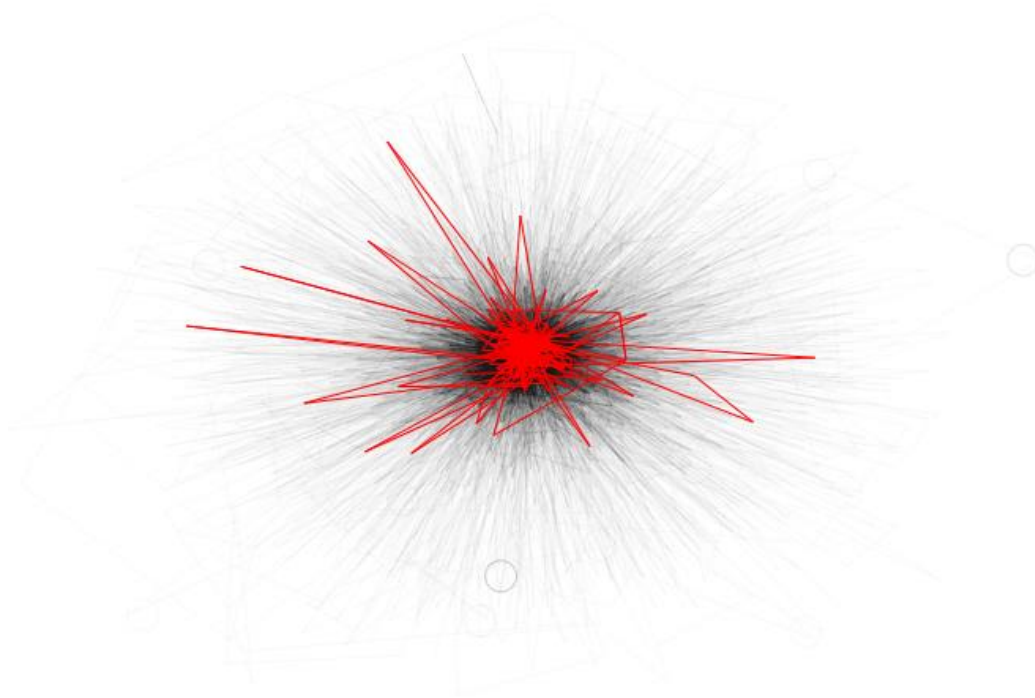


Figure 11. Structure of one Child Speech text

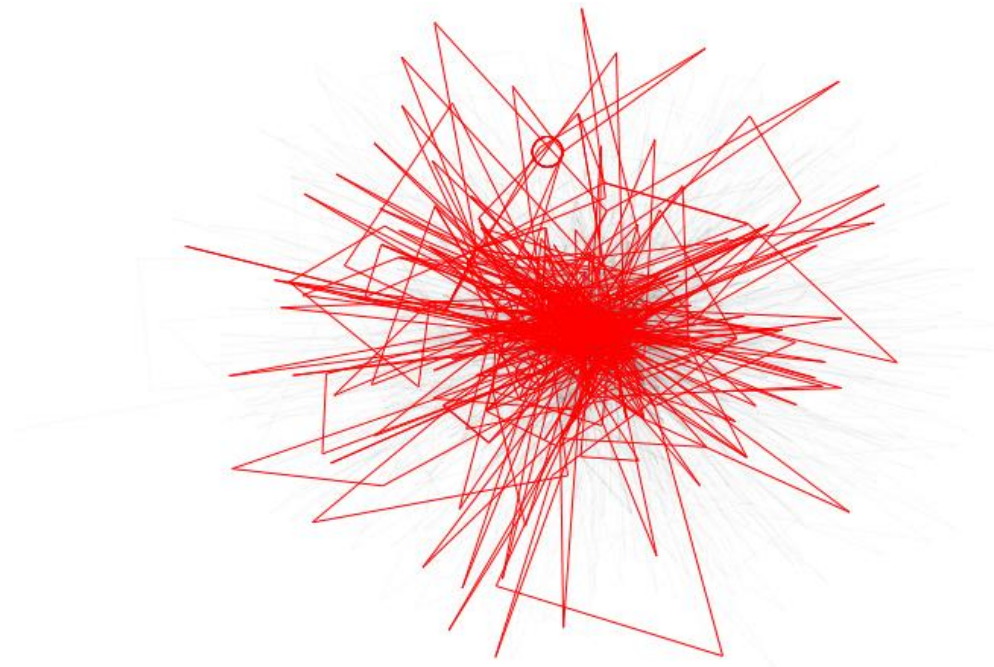


Figure 12. Structure of one Child-Directed Speech text

The graphs generated with the Fruchterman-Reingold algorithm show that we should expect more similarity between children than between adults. The high frequency bigrams are frequent for children not only because each child uses the same bigrams over and over again, but because they all do. This was also shown in Figure 19, where the dense red area of the one child's most frequent bigrams overlaid a dense black area representing the rest of the children's most frequent bigrams. The high number of infrequent bigrams from adults is so high because other adults do not use these different bigrams: there is more uniqueness in each narrative for the adults. In the next section, I present a measure to compare the number of (in)frequent words in each text as one way to quantitatively show the difference between child and adult narratives.

3.2.3. Measures of Language Proficiency

Here I show two measures of general language proficiency, text entropy and syllable rate, and how they relate to each other for speakers in the Redford Corpus. Text entropy is a heuristic for semantic complexity of a given text, and shows level of proficiency in terms of meaning and quality of word choice. Syllable rate indicates how fast a speaker can produce language and reflects both physiological development and speech planning capabilities (Redford, 2014).

Entropy is a measure of the randomness in a set of observed data (Shannon, 1948). It is applied here to the probability distributions of the texts in the present data. A perfectly predictable distribution will have entropy of 0, meaning there is no randomness in the distribution. A perfectly random distribution will have entropy equal to $\log(N)$, that is the logarithmic transformation of the number(s) we are interested in. Language is obviously neither perfectly predictable nor perfectly random.

Entropy here is an average of the log probability of the corpora, or the various texts in the corpora. Entropy for each text is calculated as the total log probability of a text, divided by the number of words in the text, or the average log probability per word (Goldsmith, 2007; Shannon, 1948). The formula for entropy of words in a set of data can be expressed as: $-\sum_{j=1}^V pr(word_j) \log_2 pr(word_j)$. This formula for entropy has a negative sign so what is expressed is a positive log, rather than a normal log. It is a sum of the probability of words in the corpus by the log probability of each word in the corpus. Entropy is used for things like evaluating data compression algorithms (Schürmann, 2004; Schürmann and Grassberger, 1996), the complexity of a

cryptographic key (Malone and Sullivan, 2005; Massey, 1994) or for evaluating morphological complexity (del Prado Martin et al., 2004).

An empirical question that can be answered with entropy is whether a child's language use is more or less predictable (or random) than that of an adult. A second empirical question we can answer with an entropy measure is whether the similarity between different children's word probability distributions are more similar than those of adults, for a given text? The frog story narrations that make up the Redford corpus are a good data source to answer these questions. The stories cover the same topics, limiting the possible vocabulary. Therefore, I do not draw from the probability of all possible English words, but from the probability of words used by school-aged children and their caregivers to narrate frog stories (as used in the Redford Corpus), meaning probabilities are from a closed set of vocabulary choices.

Words can be improbable for several reasons, associated with different ends of the proficiency spectrum. A word may be highly improbable because it is associated with more advanced vocabulary, such as: *absentminded*, *frustration*, *uncurled* or *pranced*, all of which occur only once in the Redford Corpus. Or a word may be highly improbable because it is a nonce word, like a sound effect, which is associated with a less advanced vocabulary level, such as: *arr*, *woosh*, or *zoom*, which also occur only once in the Redford Corpus. Although it should be noted that some sound effects, like *kerplunk*, occur twelve times. Finally, errors may be highly improbable, which is also associated with lower vocabulary levels, like *sitted* or *stucked* which occur once and *stinged* which occurs six times in the Redford Corpus. Therefore, it is possible that both a text replete with errors and a text composed of advanced vocabulary words may have high entropy.

As seen in Figures 13 and 14, entropy goes up as children get older ($R^2 = 0.18, p < 0.01$) and that parents' narrative entropy also increases as children get older ($R^2 = 0.13, p < 0.01$). This indicates that caregivers are adapting their speech to their children.

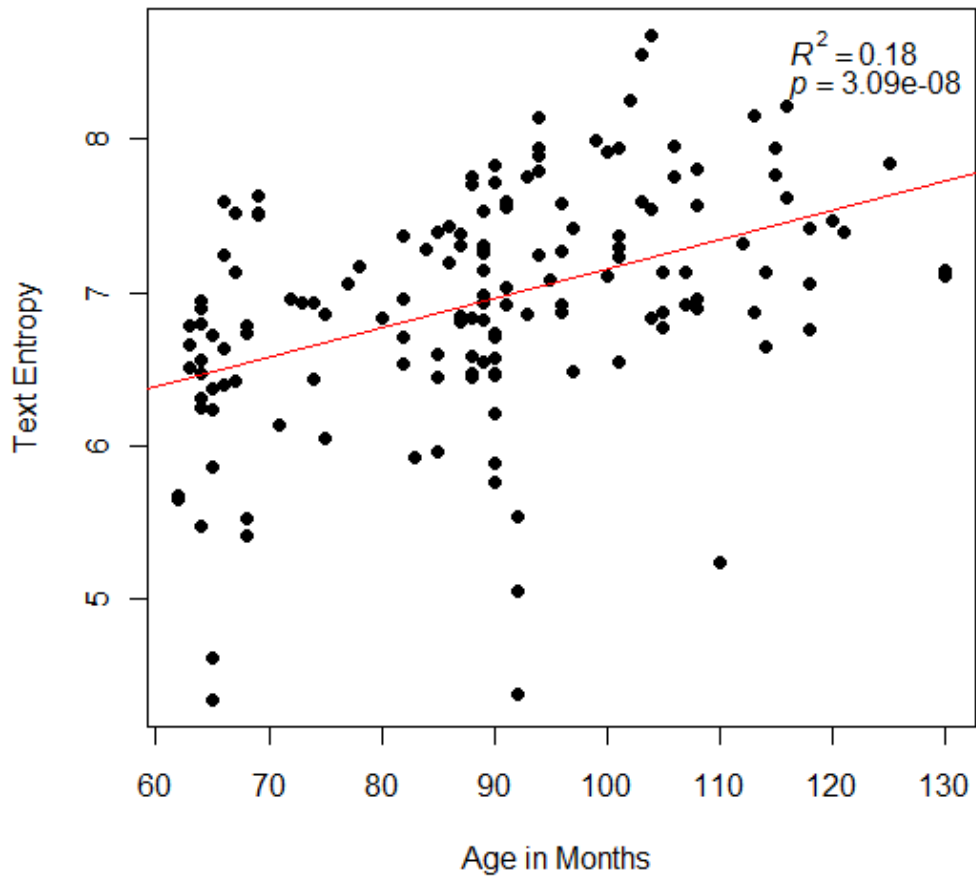
Caregivers use more familiar and frequent words in the narrative when they are speaking to younger children than to older children.

Although narrative content is adapted for child age, speech (syllable) rate is not, as seen in Figure 15. In Figure 15, we see that the speech rate of caregivers does not depend on the age of the child to whom they are speaking ($R^2 = -0.02, p = 0.75$). Syllable rate here is measured as the average time in milliseconds that it takes for a speaker to produce one syllable. Syllable rates were calculated for each utterance in a text. Pauses of over 150 milliseconds are excluded. There are fewer data points examined in Figures 15 and 16 than in Figures 13 and 14 as entropy rates were calculated for all transcribed narratives and syllable rates were calculated only for narratives that were used for the phonetic analyses (cf. 3.4 – 3.11).

Children, however, increase their speech rate over time dramatically, as seen in Figure 16 ($R^2 = 0.22, p < 0.01$).⁶ As children get older, their motor control is better and they are able to speech more quickly (Redford, 2014). For children, increased speech rate is a sign of fluency and better language skills. In Figure 17, we see that entropy and syllable rate are correlated for children ($R^2 = 0.15, p < 0.01$), showing that children who can speak faster are also children that use higher-level vocabulary words.

⁶ One outlier was excluded. One child who was 110 months old had a syllable rate of over 500 milliseconds per syllable.

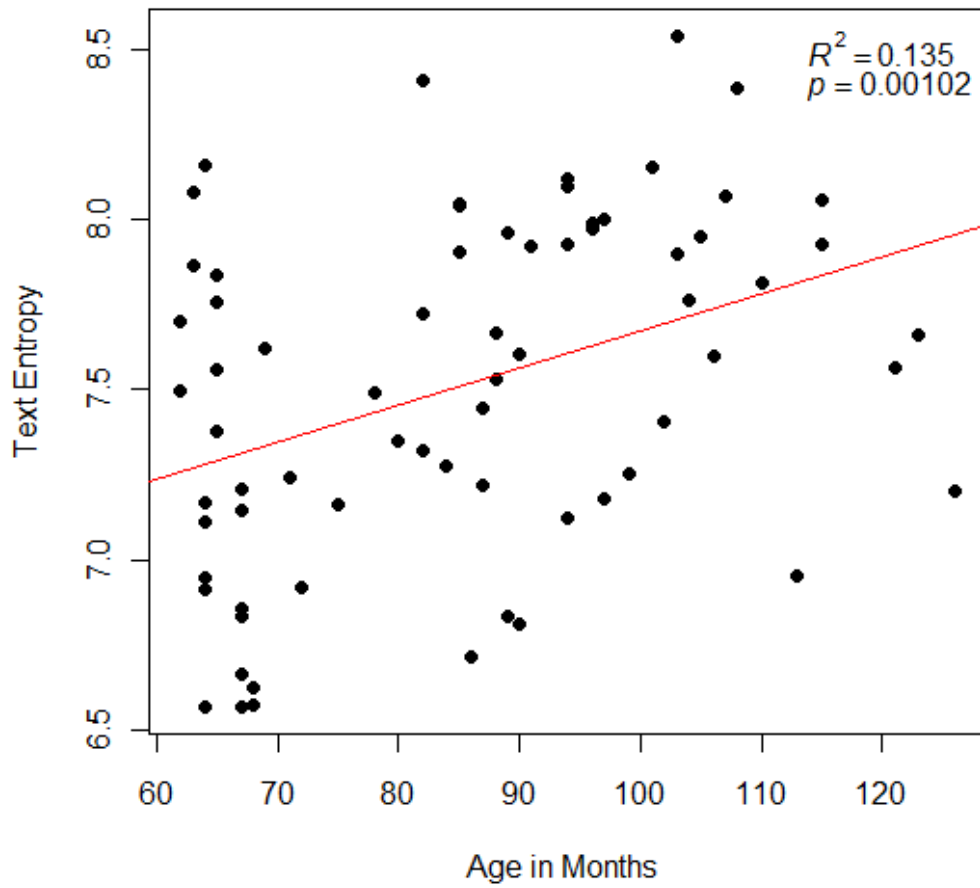
Text Entropy by Age



Note: Adjusted R-squared. Raw R-squared is 0.4305

Figure 13. Text Entropy by Age in Children's Speech

Text Entropy for Caregivers by Child Age



Note: Adjusted R-squared. Raw R-squared is 0.4305

Figure 14. Text Entropy by Age in Caregivers' Speech

Caregiver Syllable Rate by Child Age

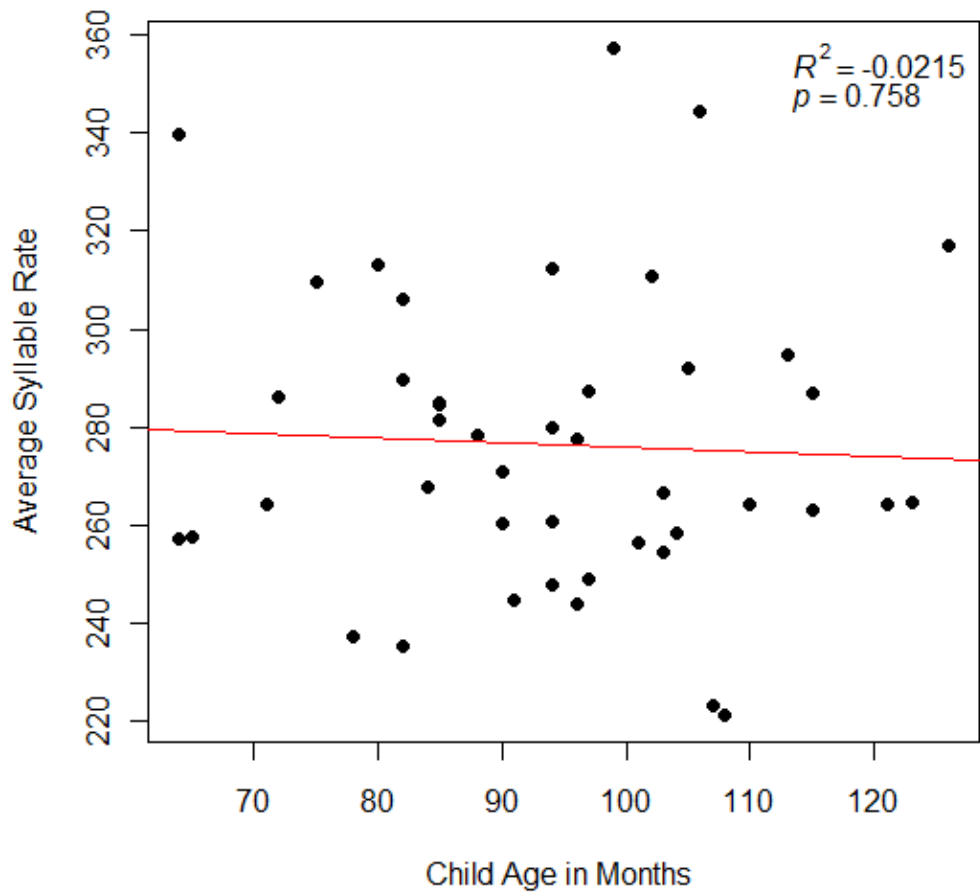
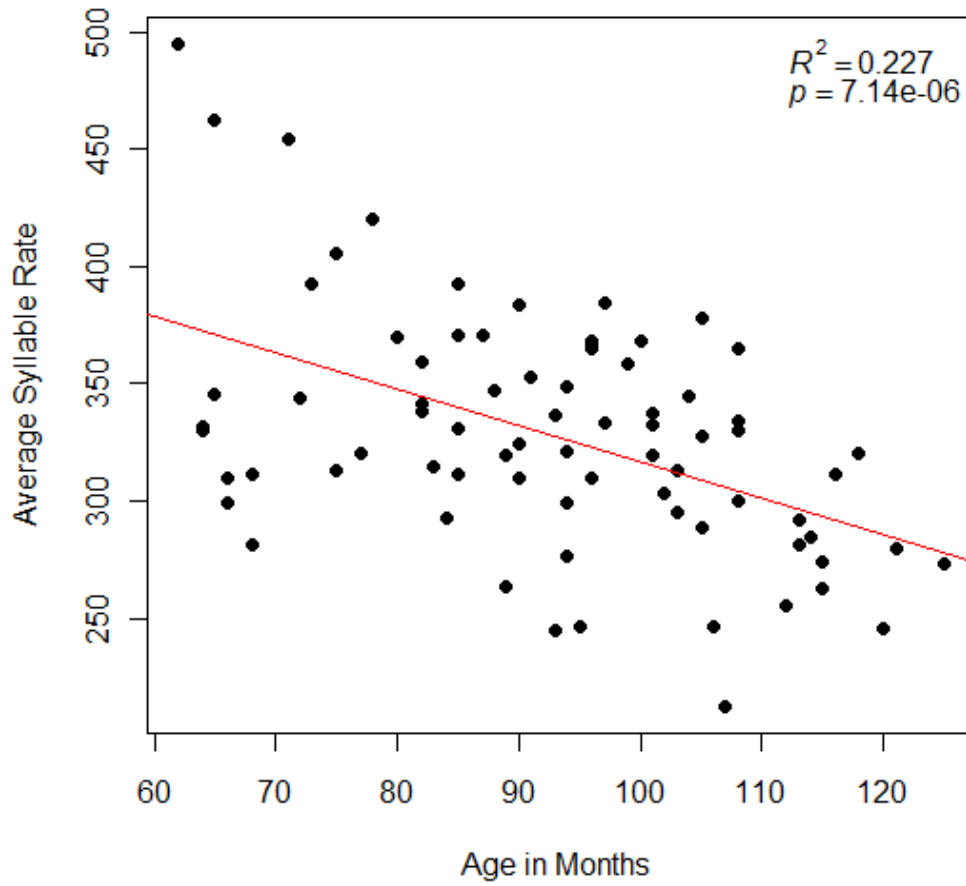


Figure 15. Syllable Rate by Child Age in Caregiver's Speech

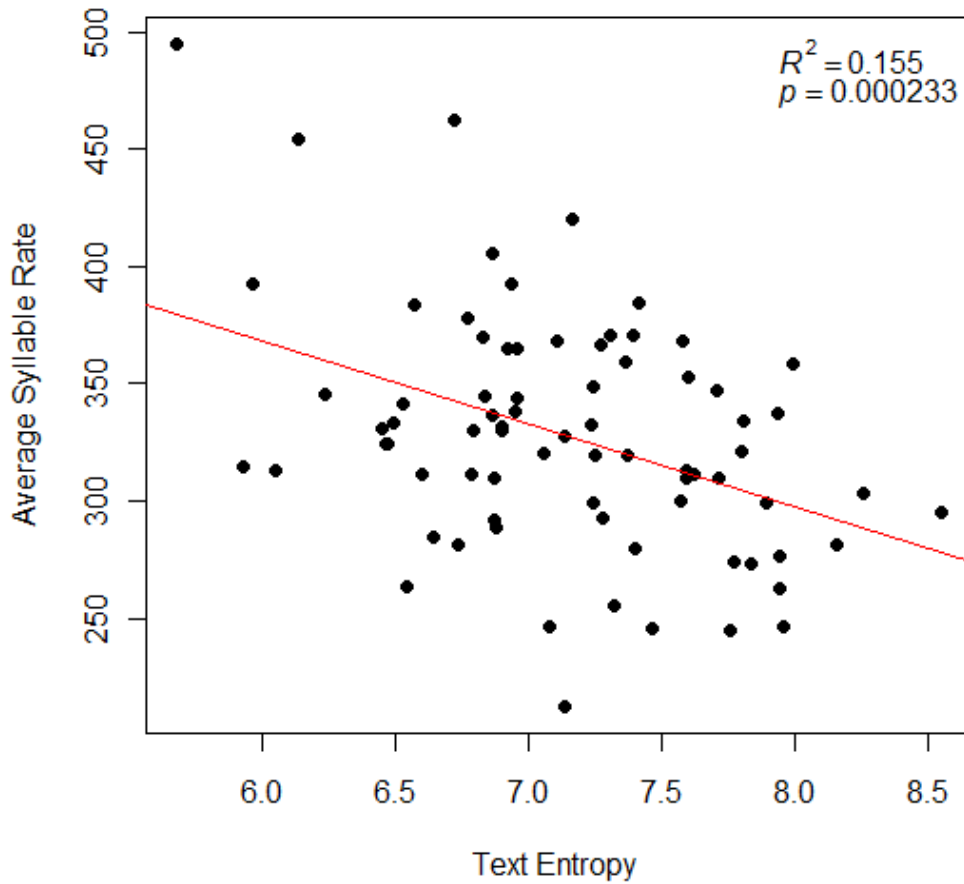
Average Syllable Rate by Child Age



Note: Adjusted R-squared. Raw R-squared is -0.48

Figure 16. Syllable Rate by Child Age in Children's Speech

Text Entropy by Syllable Rate in Child Speech



Note: Raw R-squared, Adjusted R-squared is -0.4077

Figure 17. Syllable Rate by Entropy in Children's Speech

Although child age and caregiver syllable rate are not correlated, entropy and syllable rates are correlated in caregiver speech ($R^2 = 0.236$, $p < 0.001$), as seen in Figure 18. This result indicates that caregivers who are using higher level vocabulary with their children are also speaking faster to those children, and that this is (at least somewhat) independent of the age of the child.

Text Entropy by Syllable Rate in Caregiver Speech

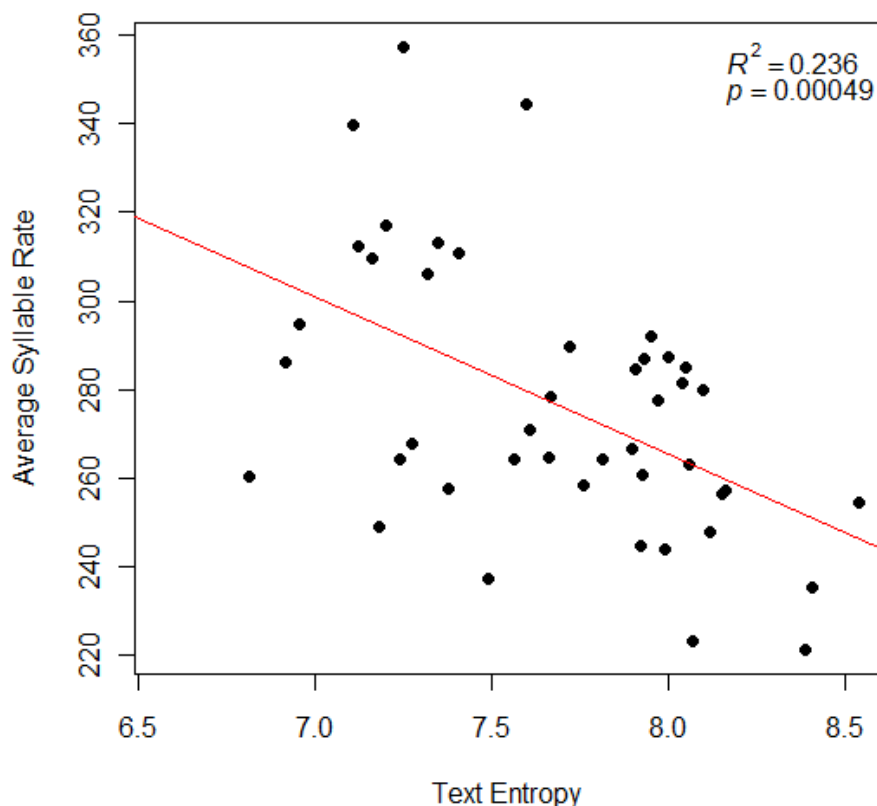


Figure 18. Syllable Rate by Entropy in Caregiver’s Speech

Figures 19 and 20 examine the relationship in syllable rates and text entropy within child-caregiver dyads. A dyad represents one time point, therefore a child-caregiver pair who participated in the study for three years in a row will have three data points in the figures below, and a child-caregiver pair who participated in the study only once will be represented by one data point. Figure 19 presents the correlation of text entropies between caregivers and their children, which is significant ($R^2 = 0.0932$, $p < 0.001$). Figure 20 presents the correlation of syllable rates between caregivers and their children, which is not significant ($R^2 = -0.023$, $p = 0.795$).

Child-Caregiver Dyad Text Entropy Correlation

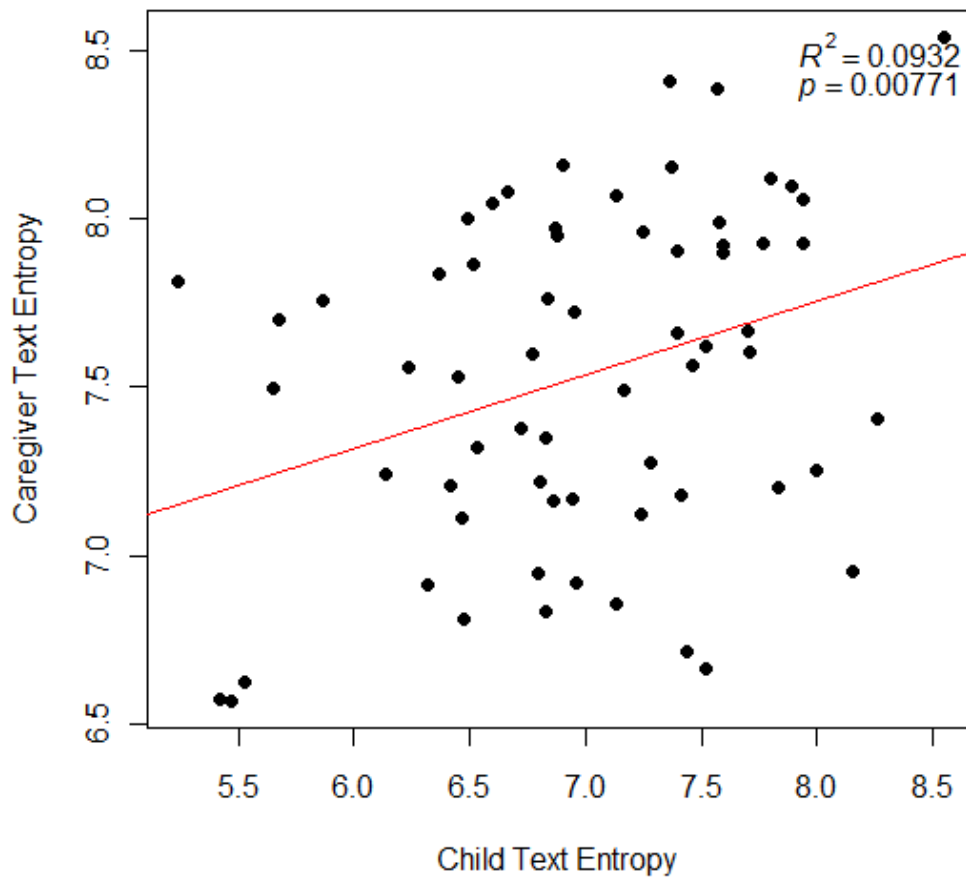


Figure 19. Caregiver-Child Dyad Text Entropy Correlation

Child-Caregiver Dyad Syllable Rate Correlation

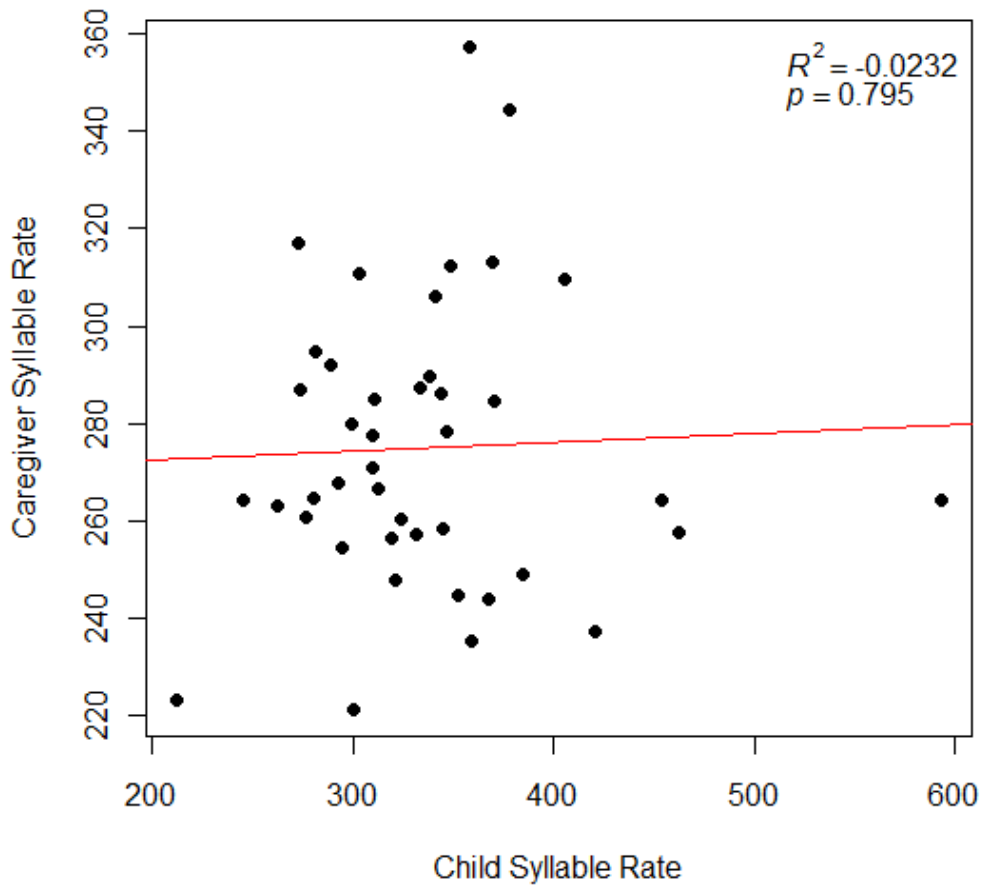


Figure 20. Caregiver-Child Dyad Syllable Rate Correlation

These figures show us that children are progressing in two clear ways (syllable rate and entropy) during the ages of 5-10. These figures also show that although parents accommodate their children in terms of content, they do not accommodate their children in terms of speech rate. The listener accommodation that the caregivers engage in does not extend to slower speech rate once their children are school-aged, despite it being a feature of CDS with younger children. The accommodation through vocabulary choice shows that caregivers are modeling their children's vocabularies and adapting their own to match. The higher rates of entropy associated with adult narratives over child

narratives could be due to imperfect modeling of their children's vocabularies. An alternate explanation is that caregivers model their children's vocabularies but nonetheless use some high level vocabulary in addition to encourage better vocabulary comprehension, and later production, in their children.

3.3. Methodology

3.3.1. Dependent Variables

3.3.1.1. Normalized Word Duration

For each target word, duration was measured in milliseconds, and this measure was normalized by the duration of the phrase and then number of syllables it contained. The software program Praat (Boersma and Weenink, 2014) was used to analyze the sound files associated with the transcripts. Two Praat scripts (Kendall, 2009a; Kendall, 2009b) were adapted to measure the duration of each target of interest. The word durations were normalized by dividing their lengths by the syllable rate of the utterance they were a part of. An utterance was defined as a string of words with no pauses. A pause was defined as a silence longer than 150 milliseconds (following Redford, 2013; 2014). The number of syllables in an utterance was obtained by using a syllable counter function (Kendall, 2011) in *R* (R Development Core Team, 2011). As an example, speaker 1016-2 produced the utterance *that was right there*. This utterance has four syllables. This particular utterance is 1191.31 msec. long, giving a syllable rate of 297.83 msec. per syllable. The word *was* is 391.71 msec. in duration. The normalized duration then is 1.315: token duration/(utterance duration/syllables in utterance). A long

normalized duration will be over 1, and a short normalized duration will be less than 1 and if the target word is exactly as long as the average syllable in the utterance, it will be 1. Normalized durations over 4 were excluded (only 1 case). The ranges of normalized durations are presented in Table 2 for each group. The word forms for the particular reported values are included in parentheses.

Table 2. Ranges of Normalized Durations by Speaker Group

Group	Shortest Normalized Duration	Longest Normalized Duration	<i>n</i>
Caregivers	0.2215 (<i>is</i>)	2.3893 (<i>is</i>)	619
Children (all)	0.21079 (<i>were</i>)	2.1425 (<i>had</i>)	755
Kindergartners & 1 st graders ⁷	0.3478 (<i>were</i>)	2.1425 (<i>had</i>)	137
2 nd graders	0.2486 (<i>were</i>)	1.9895 (<i>is</i>)	213
3 rd graders	0.21079 (<i>were</i>)	2.1025 (<i>was</i>)	293
4 th & 5 th graders	0.3019 (<i>was</i>)	1.8709 (<i>was</i>)	112

3.3.1.2. Contraction

Several words examined in the present study have the possibility of phonological reduction (contraction) in addition to phonetic reduction (shortening of duration with no reduction in length, as measured in segments). *Is, am, are, have, had* and *has* can shorten

⁷ Because of low token numbers from kindergartners and 5th graders, kindergartners were combined with 1st graders and 5th graders were combined with 4th graders. In the Kindergartner-1st grader group, speakers 1011, 1017, 1020, 1045, 1053 and 2009 have data points from two different years. In the 4th-5th grader group, speakers 1010 and 1015 have data points from two different years.

in duration or can be contracted to *'s*, *'m*, *'re*, *'ve*, *'d*, or *'s* respectively. All instances of *is*, *am* or *are* can contract but *have*, *had* and *has* only contract as perfect auxiliaries. Only inflections of *BE* occur frequently enough in the Redford Corpus to statistically analyze the contraction distribution of the word forms. Analyses for the contraction of *BE* are found in Section 3.13 and 3.14. Contraction distributions for *HAVE* are described in the same sections, but not statistically tested due to low occurrence, as can be seen in Table 3 below.

Table 3. Proportion of contraction for *is*, *am*, *are*, *have*, *had* and *has* by corpus

Word	Child Speech	Caregiver Speech
<i>is-'s</i>	357/527	365/504
<i>are-'re</i>	90/138	121/255
<i>am-'m</i>	13/13	49/52
<i>have-'ve</i>	2/3	11/19
<i>had-'d</i>	1/20	2/21
<i>has-'s</i>	2/4	16/25

3.3.2. Independent Variables

3.3.2.1. Probability and Construction Variables

3.3.2.1.1. Probability

Several probability measures were used as independent variables. Two main kinds of probability were measured: joint and transitional. Two types of frequencies were used to calculate the probabilities: frequencies from the COCA and frequencies from the Redford Corpus. This two-by-two structure resulted in four probability measures. Joint

probability is defined as the probability that two words occur together. This measure is also called string frequency (Krug, 1998). It is calculated by taking the amount of times that a string of two words (bigram) occurs and dividing it by the overall number of words in the corpus. However, joint probability can be high because two words are frequent without necessarily being predictive of each other, such as the string *of the*. Transitional probability (Saffran et al., 1996; Bush, 1999) is then used to measure predictability of a word, given the frequency of the other word in the bigram. In this way, it can be high when a pair of words is infrequent, but one word is predictable from the other every time it occurs. This is also called conditional probability (Bell et al., 2003; Manning and Schütze, 1999). It is calculated by taking number of times a string of two words occurs together and dividing it by the number of times one of those words occurs. Following common practice (Bell et al., 2009; Goldsmith, 2007) I have logarithmically transformed the probabilities because their distributions are highly skewed. The formulas I used for calculating joint probability are $\log((w+w_1)/N)$ for following joint probability or $\log((w_{-1}+w)/N)$ for preceding joint probability where N is the number of words in the corpus, w is the target word, w_1 is the word that follows the target words, and w_{-1} is the word that precedes the target word. The formulas I used for calculating transitional probability are $\log((w+w_1)/N) - \log(w_1/N)$ for following transitional probability or $\log((w_{-1}+w)/N) - \log(w_{-1}/N)$ for preceding transitional probability.

Frequencies from both the Redford Corpus and the COCA were used to calculate the probabilities. When a bigram in the Redford Corpus did not occur in the COCA, a value of 1 was used for its frequency. For many common bigrams in the COCA, logged probabilities were similar based on either Redford Corpus or COCA frequencies. The

joint probability of *is a* is $\log(105/78,000) = -2.87$ using frequencies from the Redford Corpus and $\log(592,300/450,000,000) = -2.88$ using frequencies from the COCA. The transitional probability of *is a* is $\log(105/78,000) - \log(2,000/78,000) = -1.28$ using frequencies from the Redford Corpus and $\log(592,300/450,000,000) - \log(9,907,180/450,000,000) = -1.22$ using frequencies from the COCA. However, for some bigrams in the Redford Corpus, especially ones involving frogs, turtles, etc., the joint probabilities were quite different, although transitional probabilities were still similar. The joint probability of *turtle was* is $\log(33/78,000) = -3.37$ using frequencies from the Redford Corpus and $\log(39/450,000,000) = -7.06$ using frequencies from the COCA. The forward transitional probability of *turtle was* is $\log(33/78,000) - \log(307/78,000) = -0.97$ using frequencies from the Redford Corpus and $\log(39/450,000,000) - \log(2,921/450,000,000) = -1.22$ using frequencies from the COCA.

Target words that are more predictable, as measured using transitional probability, should be shorter in duration (Bell et al., 2003). Words with a high preceding joint probability should also be shorter, however words with a high following joint probability may actually be longer, which “to some extent counterbalances the shortening effect of the conditional probability” (Bell et al., 2003: 1016).

3.3.2.1.2. Construction

For *BE*, constructions were coded as copula, future, passive and progressive. For *HAVE*, constructions were coded as modal, possessive and perfect. Construction will be investigated to see if there are any construction-specific effects of reduction beyond the specific following joint or transitional probability. For *BE*, the copula construction is the

most frequent, oldest historically and first to be acquired construction. For *HAVE*, the possessive construction is the most frequent, oldest historically and first to be acquired construction and *HAVE* in the possessive construction is a content word and should be less subject to reduction. The *HAVE* in the perfect construction is clearly a function word and should be subject to more reduction. The *HAVE* in the modal construction is semi-lexical, semi-grammatical. Here I investigate whether modal *HAVE* behaves more like its content word or function word homonym.

3.3.2.2. Speaker Variables

3.3.2.2.1. Age

The age range of children in the present study was [5;5-10] or [65-120] months. Their ages in months were included as a variable in the analyses of child speech. All adults were considered “adults” as a homogenous group, and their ages were not part of the analyses. In the present study, I am not interested in the developmental trajectory of adults which should be much shallower than the developmental trajectory of children. Additionally, I did not have age information for most of the adult speakers in my study. Age of the child listener is included as a variable for caregiver speech to investigate the accommodations made by caregivers over time.

3.3.2.2.2. Gender

Two gender variables were coded, one for children and one for parents. I investigated whether either speaker or interlocutor gender had an effect on reduction. More differences are expected for adults by gender than children by gender, as mothers

seem to show more aspects of child-directed speech than fathers (Fernald et al., 1989; Foulkes et al., 2005; Reese et al., 1996; Rondal, 1980; Snow, 1995)

3.3.2.2.3. Speaker

Speaker is included as a random effect in the mixed effects regression models. In the random forest analyses, two kinds of speaker variables are investigated for importance: Speaker (Gross) and Speaker (by text). Many speakers (but not all) in my subset of data from Redford Corpus produced 2 or 3 narrations, over the course of the 3 years of data collection. When using speaker as a random effect, then, I could use the “speaker” from each particular narration (something like person-1003-year_1 and person-1003-year_2 and person-1003-year_3), but this would mean that the same actual person is represented by three levels in the random effect structure. Alternatively, I could use “Speaker” to represent a particular individual for each narration they contributed (person-1003 for each text they produced) resulting in fewer levels in the random effects structure. In the former, there are fewer tokens per level of the factor and in the latter, there are more tokens per level. The former risks over-fitting due to fewer cases per level and the latter risks attributing a lot of variation to the random effects because a participant’s behavior may change dramatically over three years, particularly if that participant is a child. Therefore, both kinds of “Speaker” were included as factors in random forest variable importance rankings to investigate which grouping better accounted for the data and which would be used as the random effect term in the mixed effects models.

3.3.2.2.4. Syllable and Contraction Rates

Although word durations were normalized by the utterance they occurred in, overall syllable and contraction rates were included as control variables in the analyses. An average syllable rate was calculated for each speaker by averaging the syllable rate of each utterance in the corpus, whether or not it included a target word of interest. The measure is operationalized as the average length (in milliseconds) it takes for a speaker to produce one syllable.

Additionally, I calculated a contraction rate for each speaker. The contraction rate for a speaker is the proportion of times that they contracted a word when it was possible to contract. Possible contractions were: contractions of *is*, *am*, *are*, *would*, *did*, *does*, *will*, *not*, perfective *had*, *has*, *have* and the reductions of *going to* to *gonna*, *want to* to *wanna*, *do not (don't) know* to *dunno*, *out of* to *outta* and *got to* to *gotta*.

3.3.2.3. Control Variables

3.3.2.3.1. Quartile

Because word lengthening may increase as a function of utterance length, over the utterance, the target's position in its utterance was taken into account. Targets were coded as occurring in the first, second, third or fourth quartile of the utterance and this variable was used as a factor (non-numeric) predictor. Rather than an exact measure, a heuristic was used to calculate the rough position of a target in its utterance. Utterances were coded for number of characters and the position of the first character of the target word in this total character count from the utterance was used to determine the word's quartile. Quartile is a control variable in the present study. I am interested in examining the effects

of probability and speaker-based variability while statistically controlling for the effects of utterance quartile.

3.3.2.3.2. Pauses and Disfluencies

Words were coded for being preceded or followed by a pause, error or disfluency. A pause was defined as 150 milliseconds or more for the Redford Corpus (following Redford, 2013, 2014). A pause delineated an utterance boundary. An error was defined as a restarting of a word or stopping in the middle of a word. Function words are highly unlikely to be preceded or followed by a pause in normal discourse (an exception is ellipsis ‘*I thought he wasn’t going but he was*’), so pauses generally indicates the presence of a disfluency. Other disfluencies coded were elements like “um,” “uh,” or, “er” as these were usually cases of a filled pause or a speaker searching for the next word. Words in disfluent contexts are likely to be longer and contain unreduced vowels (Bell et al., 2003 *inter alia*). Disfluency is a control variable in the present study. I am interested in the effects of probability and speaker-based variability that happens while controlling for the effects of neighboring disfluencies.

3.3.2.3.3. Stress Context

Targets were coded for whether they were preceded or followed by a stressed syllable, as speakers may be more likely to reduce after a stressed syllable (Echols and Newport, 1992; Davis et al., 2000; Jusczyk et al., 1999; Young, 1991). Stress context is also a control variable in the present study.

3.3.2.3.4. Phonological Context

Targets were coded for whether they were preceded or followed by a vowel or a consonant, as some studies have shown that contraction occurs more frequently after

vowels than consonants (Labov, 1969; MacKenzie, 2012; McElhinny, 1993), although the presence of vowels before a target and the presence of a pronoun before the target, which also lead to shortening or contraction, are highly correlated.

3.3.2.3.5. Subject Noun Phrase Type

Research has shown that *BE* is more likely to contract after a pronoun, particularly after a personal pronoun, than a full noun phrase (Barth and Kapatsinski, Forthcoming; Krug, 1998; Labov, 1969). To determine if this effect is due to the higher transitional (or joint) probability of pronouns + *BE* or due to word class, Subject Noun Phrase Type is included as a control variable in the analyses below. This factor has three levels: personal pronoun (*I, you, he, she, it, they, we*) v. non-personal pronoun (e. g. *what, who, there, here*) v. nominal noun phrase (e. g. *the frog, the boy, a baby*).

3.3.2.3.6. Priming

I investigated effect of priming in two senses: contraction priming contraction and shortening resulting from repeating a word within a short time span, as several studies have shown that speakers are more likely to use a particular word form or structure if they have previously used it (Barth and Kapatsinski, In Press; Cameron and Flores-Ferrán, 2004; Poplack, 1980; Scherre and Naro, 1991; Torres Cacoullos and Travis, 2013; Travis, 2007) and that repetition of a word makes it more accessible (e.g. Forster and Davis, 1984) which could lead to shortening/reduction (e.g. Baker & Bradlow, 2009; Fowler, 1988; Fowler & Housum, 1987; Galati & Brennan, 2010).

If there were any instances of contraction or reduction (as listed above in the section on contraction rate), this was considered potential priming for further contraction. The 10 words before each target were examined for an instance of contraction. This

variable was binary, if there were one or more instances of contraction, it was coded as contracted (1), otherwise it was coded as uncontracted (0).

When a speaker produces a word they have recently produced, they may expect that that information is more accessible to their listener, leading to shorter productions, as less listener accommodation is needed. Or a recently produced word may be more accessible for a speaker, resulting in easier processing, which in turn results in a shorter production. Therefore, for each target word, the preceding ten words were examined for instances of the lemma of the target.

For *is*, *am*, *are*, *was* and *were*, the previous ten words were examined for contracted variants of *is*, *am* and *are*, non-contracted variants or other inflections of *BE*, like *be*, *been*, *being*, *was*, and *were*. Barth and Kapatsinski (In Press) show that contracted variants are likely to precede contracted targets and non-contracted variants are likely to precede non-contracted targets, but that other inflections have no effect on the likelihood of the target word being contracted. It is an open question in this research if contracted v. non-contracted forms have an effect on duration measures of *is*, *was* and *were*. This variable has five levels: (1) contracted, (2) non-contracted, (3) non-contractible (4) not present and (5) both contracted and not contracted present.

For *had*, *has*, and *have*, the previous ten words were examined for contracted variants of the perfective auxiliary or any inflection of *HAVE*. There was only one case of *having* occurring within ten words of a target word, so this case was collapsed with the presence of *had*, *has* or *have*. There were only four cases of a contracted variant of *HAVE* occurring before a target word, so these cases were collapsed with the presence of *had*, *has* or *have*. This variable has two levels: (1) lemma present and (2) lemma not present.

All the priming variables are also control variables in this study. I am interested in the effects of reduction that are not due simply to priming.

3.3.3. Random Forests

Random forests are an analysis based on a type of recursive partitioning analysis called classification trees, or *ctrees* done in the *party()* package (Horton et al. 2006a, Horton et al. 2006b, Strobl et al. 2007, Strobl et al. 2008) in *R*. This kind of model determines which variable at which level(s) makes the best binary split of the data. Then after the first split, the classification tree determines which variables, or levels of a variable, makes the best split of the remaining cases under each node. This continues until a stopping criterion is reached. Because each set of data under a node is looked at anew, the same variable can be used again in a lower level of the tree, with different levels. This feature of the classification makes it useful to explore nonmonotonic relationships. *Party* classification trees also provide *p* values for each node indicating whether the group difference indicated by the split was significant. The algorithm for creating the tree model will not necessarily use all independent variables listed in the model specification. If there are IVs that would not make a significant split in the data, they go unused.

Figure 21 shows an example classification tree. In a random forest analysis, many classification trees are computed based on different subsections of the data and subsets of IV levels. Figure 22 shows a slightly different classification tree based on all but 30 of the cases of the data used to create the classification tree in Figure 21.

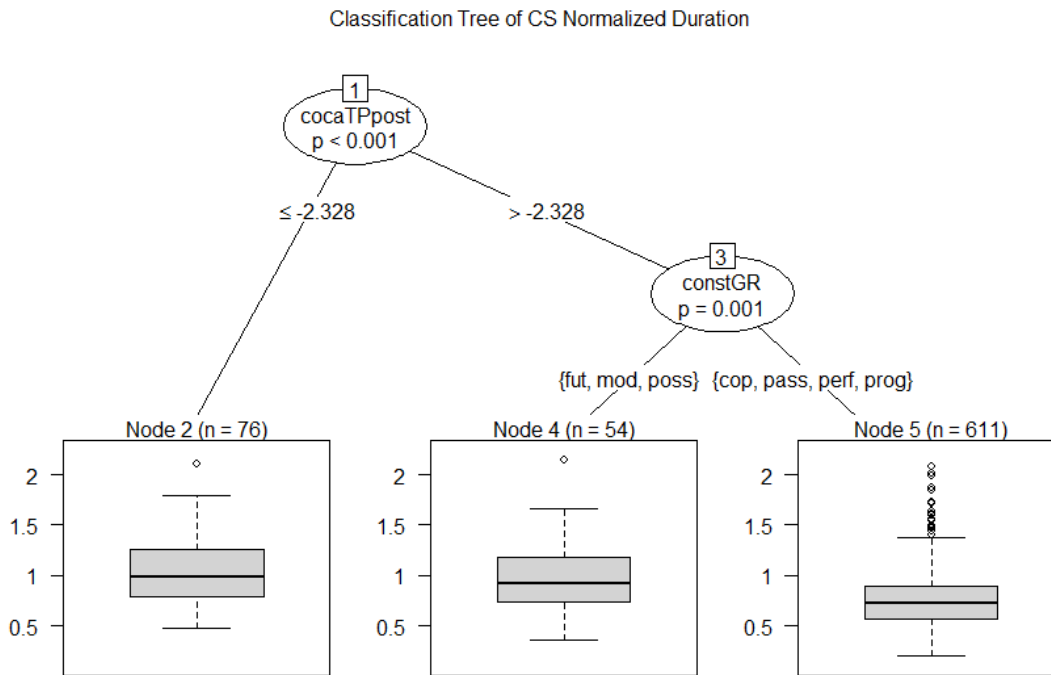


Figure 21. Example Classification Tree

When many classification trees are averaged, factors can be ranked by their importance, determined by which factors most often make a significant split in the data. Random forests are very useful for determining which of several collinear predictors is the best predictor for the data (Schneider, 2014; Tagliamonte and Baayen, 2012) because the resampling of data involved in generating a random forest reduces the possibility that a variable will be considered important by overfitting the noise in the data, while resampling of predictor levels helps deal with collinearity. Slightly different data subsets will result in potentially different groupings performing well (sometimes preceding phonological context, sometimes preceding pronoun vs. nominal, sometimes preceding JP, sometimes preceding TP, etc.) but after looking at many cases (a few thousand), one will generally be the best factor for predicting the DV. This one factor will be ranked higher than the other. For each of the dependent variables, I rank here potentially

influential independent variables by importance using random forest analyses. Several of the potentially influential factors are highly multicollinear, making them inappropriate to combine in a regression analysis. A random forest analysis using `cforest()` allows for a comparison of collinear factors, and the factors are ranked by importance using `varimp()`.

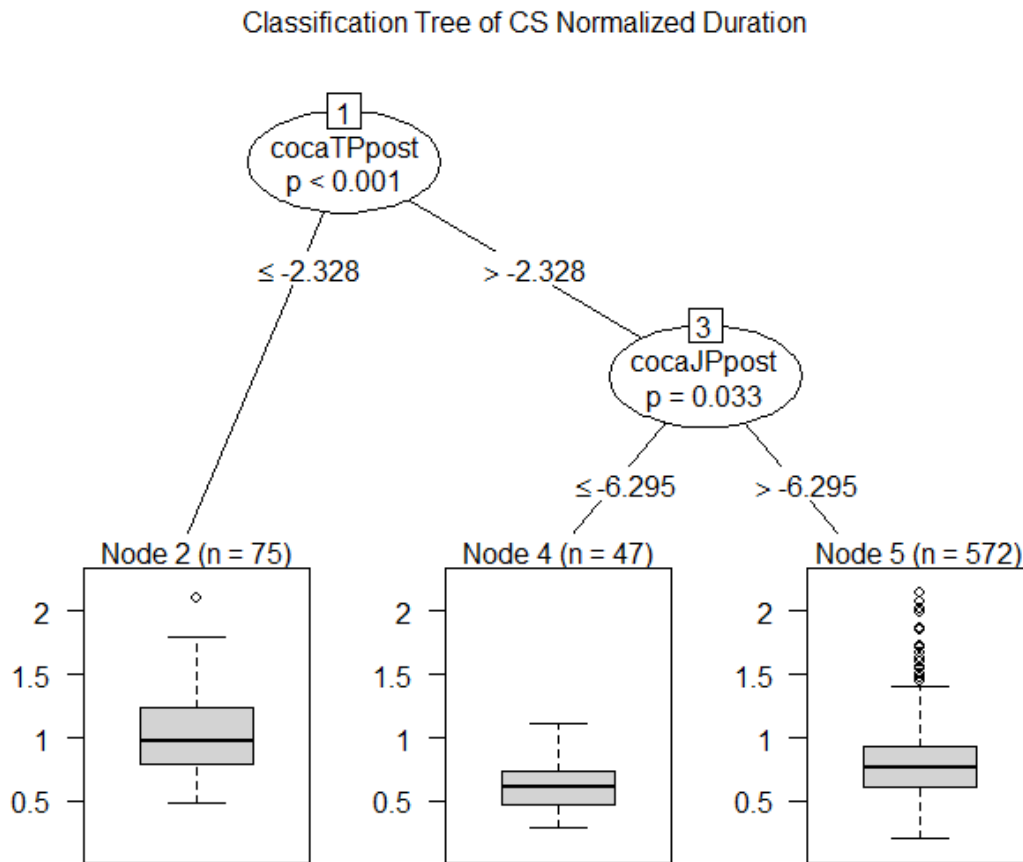


Figure 22. Example Classification Tree on New Data Subset.

In sections 3.4-3.14, I examine the potentially influential factors in the reduction of grammatical words *am*, *are*, *had*, *has*, *have*, *is*, and *was* in both caregiver and child speech. I present variable importance rankings for each collinear group of factors:

preceding context factors, following context factors and speaker-specific factors. Results from the function `varimp()` where a factor is ranked above a particular threshold (the absolute value of the ranking worst performing factor) are considered significant (Shih and Grafmiller, 2011). If a factor is ranked as important in the random forest analyses, it can potentially be included in a regression model for significance testing. The random forest analyses here combine all grammatical words. The best performing factors from each group (preceding context, following context and speaker-specific) are then tested for correlations in each subset of data for model testing *had-has-have*, *is*, and *was*. The best performing factor or factors from the group random forest analyses that are not correlated for the specific data subset are then selected for the regression models.

Because a word's phonemic structure also influences its length, "lemma" is included as a potentially influential factor as well in the random forest analyses. It is grouped with the speaker-specific factors, as a word-specific factor, although it should not be collinear with any of the speaker-specific factors.

For child speech, random forest analyses are done for the children as a group, but as there is expected longitudinal development in reduction behavior, random forest analyses are also presented for each grade of children, kindergarten through 5th grade.

3.3.4. Multimodel Inferencing

Based on the random forest analyses, regression models are built using the best performing, non-collinear factors. Collinearity is tested for each specific subset of the data for which a regression model is built using *mixed.cor* in the *psych()* package producing Pearson correlation coefficients (Revelle, 2014). Regression models are

presented for all words together in caregiver speech, models for each word of interest in caregiver speech when word specific results differ from the all-word model, all words together in child speech with an interaction for child age and models for each word of interest in child speech when word specific results differ from the all-word model. The kind of regression analysis done here is multimodel inferencing (Barth and Kapatsinski, In Press; Burnham and Anderson, 2002; Kuperman and Bresnan, 2012). Multimodel inferencing builds all possible models out of a given set of predictors, up to a specified maximal model, as well as a null model with no predictors. If three factors were of interest, there would be three models with each factor alone, two models with each combination of two factors, one model with all three factors and the null model for 10 models total, which would be ranked by their corrected Akaike Information Criterion or *AICc*. For some of the models reported below, there are many possible models, so only the models with possible predictive value are reported, those that are within $\Delta 2$ in *AICc* of the best performing model, as only these have substantial empirical support based on the data (Burnham and Anderson, 2002). I also report the null model, as a way of comparing how much support the best performing models have. All models reported below substantially outperform the corresponding null model.

For each set of models, the performance of factors is averaged. Factors that perform well in many models, and especially many highly predictive models, have a high cumulative probability (CP) score, indicating that they are highly probable of being predictive. I report coefficients with shrinkage, adjusted standard error and significance values that have been punished or “shrunk.” Factors that are significant are in bold. Factors that have high cumulative probability, but are not significant are bold and grey.

Often factors that have high cumulative probability but p values more than 0.05 were significant before punishment. As the number of potential factors increases, the severity of the punishment increases. In model building, many of the models included several non-significant, non-probable factors. These are factors that random forest analyses identified as potentially contributing to predictiveness, but that were not significant when random effects, etc. were taken into account and relationships were constrained to be linear (rather than arbitrarily shaped and potentially even non-monotonic) in the regression model. Using backward-elimination, I removed factors that had a CP of less than 0.4. In model building, some of the model outputs had w values that were quite low. The w indicates the Akaike weight of a model, and the w of all models sums to one. Therefore, with many possible models w of any one model is quite low unless one model substantially outperforms all others. When the non-probable, non-significant factors were removed and the same the multimodel procedure was done, the cumulative probability values remained fairly stable, but p values for individual factors and w values for the highest ranked models went up. In the reported models below, non-probable factors have been removed but some non-significant predictors remain as they have CP scores of over 0.40. Discussion of the predictors and of the model ranking results follows each table of results.

3.4. Word Shortening in Child-Directed Speech

3.4.1. Introduction

The analyses of word shortening in child-directed speech are aimed to answer the following primary research questions:

Q1. Are the function words *HAVE* and *BE* subject to probabilistic reduction?

Q2. Do caregivers adapt their speech in one or more phonetic characteristics to school-aged children?

Q3. Do adults who contract more often also shorten grammatical words to a greater degree than those who contract less often?

Q4. Do adults differentiate function word production by construction (meaning)?

In this results section on word shortening in caregiver speech, I first present random forest analyses and then multimodel inferencing results. I present three random forest analyses for each group of potentially collinear factors: speaker (and word) based factors, preceding context factors and following context factors. Based on the rankings of those analyses, I present multimodel inferencing analyses, building models from highly ranked, non-collinear factors. During the model building process I eliminate factors with low cumulative probabilities (under 0.4) that are highly unlikely to make a positive contribution in a model. In the model building process, I also test for *BE* and *HAVE* specific factors. As it will be seen below, only *HAVE* constructions show differences from the combined analysis of all words, and so two model outputs are presented, one for all words combined and one for instances of *HAVE*.

3.4.2. Random Forest Variable Importance Rankings for Word Shortening in Child-Directed Speech

Figures 23 through 25 present the variable ranking importance for caregivers' speech. Figure 23 shows that for caregivers, the length of the particular word was important, which is expected. A word-specific effect of length shows that it is appropriate

to look for word-specific interactions analysis of normalized duration and to include word inflection as a random effect. The contraction rate of individual speakers is also ranked as important. Contraction rate is a measure of how often a speaker contracted a word given the possibility of contraction, for a given text. Those who contract grammatical words often potentially also reduce (or fail to reduce, as directionality is not specified in the random forest analyses) grammatical words phonetically more often as well. Both Speaker (Gross) and Speaker (by Text) are ranked as important, but for the caregivers, Speaker (Gross) is ranked as more important. Speakers that participated in the study more than one year, have contributed more than one story to the data. The random forest analysis shows that behavior of the particular speaker is fairly constant across years or texts. For adults, this is unsurprising; their behavior in reduction should be stable over a few years. This ranking also indicates that Speaker (Gross) is a more appropriate random effect for the regression analyses than Speaker (by Text). Utterance quartile is also ranked highly, indicating that a word's position in an utterance affects its length, presumably word lengthening in the last quartile of the utterance. This effect will be examined in more detail in model testing. There are also several factors that are rated as relatively unimportant, although they do have importance values above the threshold, and therefore will be used in the regression models after testing for correlations. The gender of the child matters to some degree, as does the age of their child. The gender of the caregiver barely reaches the threshold. The average syllable rate of the speakers per text is not ranked very highly. As the word durations have already been normalized, this is unsurprising. However, when we compare this factor to contraction rate, we see that grammatical word reduction is related to a speaker's propensity to contract other

grammatical words, rather than their speech rate. One reduction behavior influences another. It is not the case that fast talkers reduce (these) grammatical words to any higher degree than other words in their utterances.

Speaker and Word Variable Importance for CDS duration

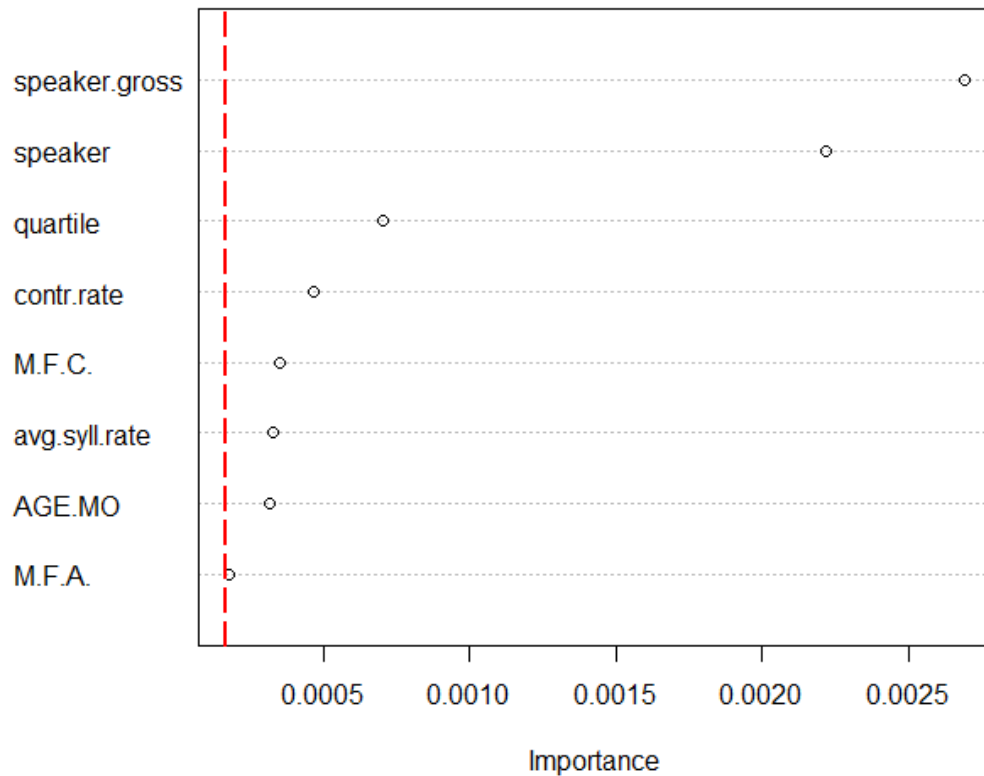


Figure 23. Speaker-specific and word-specific variable importance for CDS grammatical word duration

Figure 24 shows the importance of the preceding context variables in caregiver speech. The best performing variables are all multicollinear. The four probability variables are all ranked as important: Preceding Transitional Probability based on frequencies from the COCA, Preceding Joint Probability based on frequencies from the COCA, Preceding Transitional Probability based on frequencies from the Redford

Corpus and Preceding Joint Probability based on frequencies from the Redford Corpus. The Joint Probability measures outperform both of the Transitional Probability measures. When joint probability has a stronger influence than transitional probability, it indicates that there is a strong bond between elements in a bigram, here between the subject and the verb or auxiliary. The probabilities based on the COCA frequencies also outperform the probabilities based on the Redford Corpus frequencies. As the COCA frequencies are based on much more data, their effects should be more robust and reliable than the frequencies based on the narrative task. Additionally, probabilities based on COCA frequencies outperforming probabilities based on Redford Corpus frequencies shows that caregivers are more sensitive to global or overall word experience, rather than being more sensitive to the predictability within particular task they are performing: narrating frog stories. Whether or not the target is preceded by a stressed syllable also has an effect on duration, as does whether or not the preceding element is a pronoun or not. Both of these factors are highly collinear with the probability factors and it is likely that the effect of stress or pronouns is due to the particular preceding word and its frequency. The next three variables are priming variables. If a *HAVE*, *BE* or contraction has preceded the target recently, it is possible that that reduction would prime further reduction. We see that only a preceding *BE* has an effect that is ranked as important. There may not be enough instances of *HAVE* preceding targets for it to have an effect. We also see that recent contraction has no effect, although overall contraction rate does (cf. Figure 23). Finally, we see that a target preceded by a pause is not likely to be any shorter or longer than a target not preceded by a pause.

Preceding Context Variable Importance for CDS duration

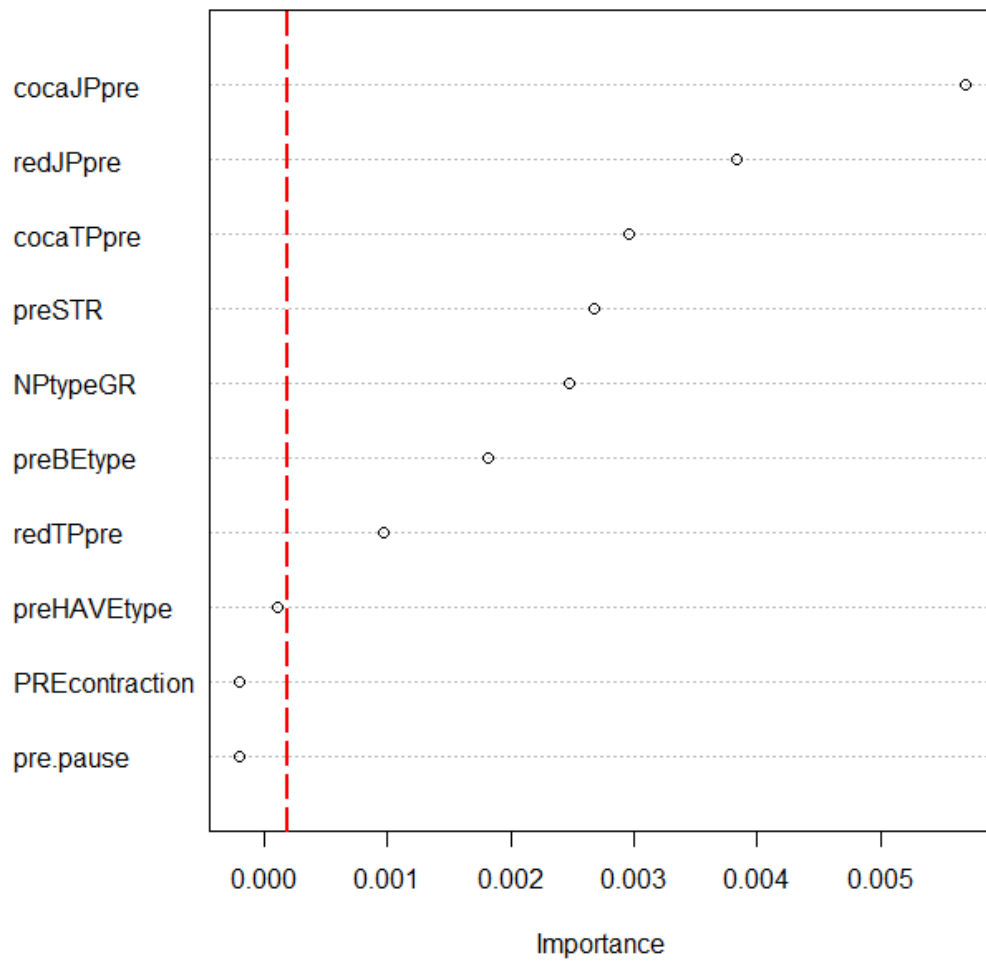


Figure 24. Preceding Context variable importance for CDS grammatical word duration

Figure 25 shows the importance of the following context variables in caregiver speech. The most important variable is whether or not there is a following pause. There are only 28 instances where a target is followed by a pause, filled pause or error, but these instances are demonstrably longer than the others. Based on my posthoc examination of the data, all of these instances occurred when a speaker was searching for a word. The longer durations then, are likely used by speakers to buy time to access the

word they are trying to find (Bell et al., 2003; Kapatsinski, 2005; Schachter et al., 1991; Schnadt, 2009). Because disfluencies occur in low probability contexts more than in high probability contexts (Beattie & Butterworth, 1979; Goldman-Eisler, 1957; Maclay & Osgood, 1959; Schachter et al., 1994; Schnadt, 2009, Tannenbaum, et al. 1965), it will be important to examine the correlations of pauses/errors and probability for each data subset before putting both factors in a regression model. The constructional meaning of the target is also important. Constructional meanings differ based on the particular words, so this factor will be examined in more depth in the regression analyses. All four probability variables are ranked above the threshold, but barely: Following Transitional Probability based on frequencies from the COCA, Following Joint Probability based on frequencies from the COCA, Following Transitional Probability based on frequencies from the Redford Corpus and Following Joint Probability based on frequencies from the Redford Corpus. Finally, we see that whether or not the following context is stressed or not has no influence on the duration of the target words, although the stress of the preceding context is ranked as important (cf. Figure 24). Following context could matter less for most of the auxiliaries as they are more likely to form a bond with the preceding word (Bybee, 2002). The exception to this is the modal *HAVE*, which can form a bond with the following context: *hafta, hasta, hadda*. As this specific case of potential bonding before an unstressed monosyllabic word (*to*) is found only in modal *HAVE*, the constructional meaning variable can account for the potential effect of the stress of *to* on the duration of *HAVE* in this circumstance. However, Bell et al. (2009) show that following context probability is more likely to have an effect on word duration than preceding context probability.

Following Context Variable Importance for CDS duration

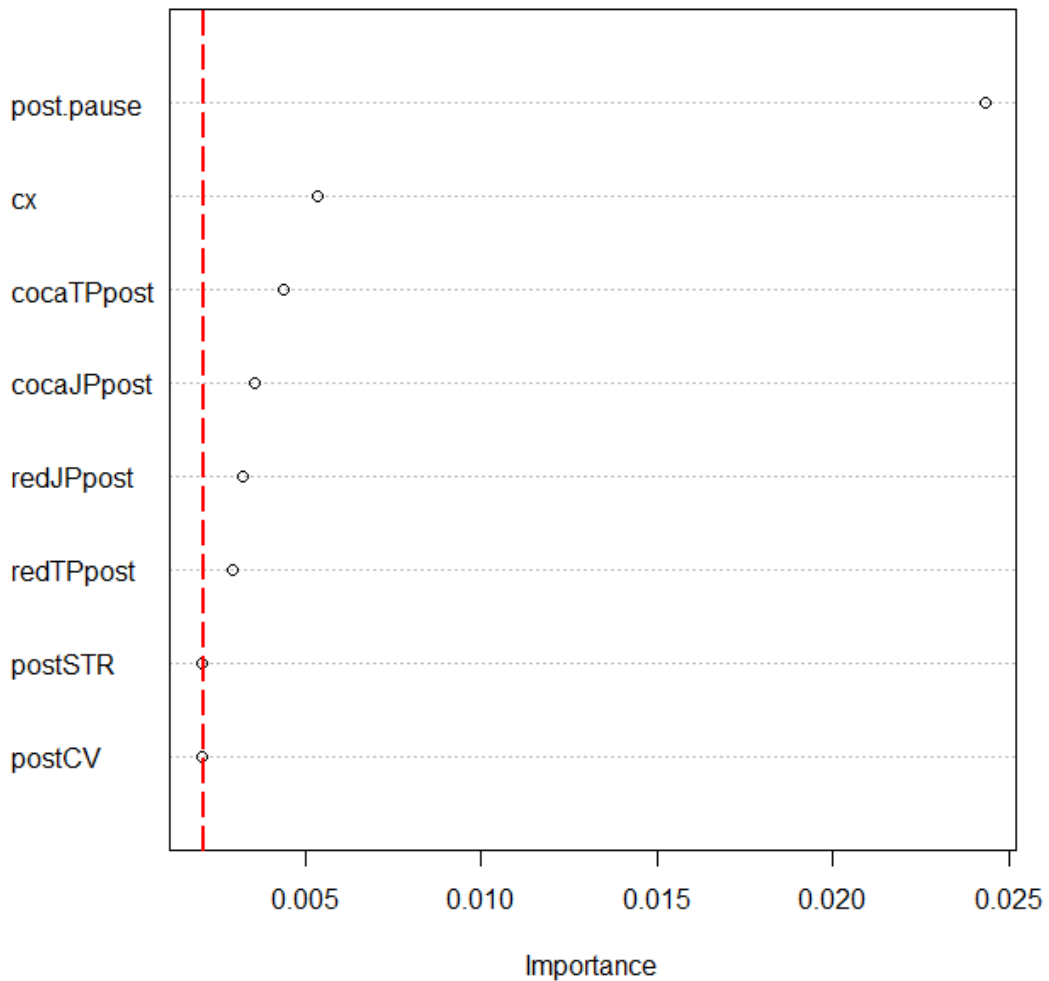


Figure 25. Following Context variable importance for CDS grammatical word duration

3.4.3. Multimodel Inferencing Analyses for Word Shortening in Child-Directed Speech

Based on the random forests, the following variables were initially selected for model testing for caregiver speech: Lemma (as a random effect), Speaker (Gross) (as a random effect), Quartile, Speaker Contraction Rate, Speaker Gender, Child Gender, Age of Child, Preceding Joint Probability based on COCA frequencies and Redford Corpus frequencies, Preceding BE Type, Preceding Stress, Construction, Following Transitional

Probability based on COCA frequencies and Following Pause. These variables were ranked as important in the random forests and do not correlate with each other (cf Table 4 below for an example of a correlation check). Preceding BE Type is only relevant for the *is* and *was* datasets, as this is a variable capturing the priming of uttering a word similar to the target word within ten words before the target. There is no theoretical reason for a preceding *BE* to prime and therefore reduce a form of *HAVE*. A model is built for examining normalized duration overall for child-directed speech and this model contains neither construction, which has different levels depending on the specific word, nor Preceding BE Type, which is not relevant for all word types, as explained above.

The first model built to compare normalized word durations for caregivers combined all word types: *had* ($n = 71$), *has* ($n = 26$), *have* ($n = 42$), *is* ($n = 127$) and *was* ($n = 356$), (total $n = 622$). Random effects for this model are Speaker (Gross) and Word. Preceding context variables were selected from the random forest analysis of preceding context variables (Figure 24) and significant, applicable variables were checked for correlation using *mixed.cor* in the *psych()* package producing Pearson correlation coefficients (Revelle, 2014). COCA preceding joint probability was highly correlated with COCA preceding transitional probability ($r = 0.66$) and Noun Phrase Type ($r = 1$), but not with Redford Corpus preceding joint probability ($r = 0.41$), nor with Preceding Stress ($r = 0.15$), which is also not highly correlated with Redford Corpus preceding joint probability ($r = 0.42$). Therefore, included in the model are COCA preceding joint probability, Redford Corpus preceding joint probability and Preceding Stress. Following Context variables were selected in the same manner, referencing the random forest analysis of following context variables in Figure 25. Following pause was not highly

correlated with any of the following probability measures (see Table 4 below), but both Redford Corpus probability measures were highly correlated (> 0.6) with one of the COCA probability measures, which ranked higher in the random forest analysis, and so were not included. As the COCA following probability measures correlated highly with each other ($r = 0.25$), they were both included in the model with Following Pause. Speaker was a random effect, but Speaker Contraction Rate, Speaker Gender, Child Age and Child Gender were checked for correlations and there were none. These four speaker-based variables were included in initially model testing. All variables were checked for interaction with specific word forms, and none were found.

Table 4. Pearson Correlation Matrix of CDS Following Context Variables for CDS Duration Analysis

	Following pause	COCA TP post	COCA JP post	Redford JP post	Redford TP post
Following pause	1				
COCA TP post	0.44	1			
COCA JP post	0.34	0.25	1		
Redford JP post	0.28	0.24	0.75	1	
Redford TP post	0.19	0.65	-0.22	0.05	1

Note: variables listed in order of random forest analysis rankings. Variables that correlate above 0.6 are in bold. Correlations calculated using the function *mixed.cor* from the {Psych} package in R (Revelle, 2014).

Variables that had a cumulative probability lower than 0.4 were considered highly unlikely to have a strong effect, and were eliminated in a backward (step-down) selection procedure⁸.

Results from the multimodel output of the caregiver duration model are presented in Tables 5 and 6 below. In Table 5, we see that five factors have high cumulative probability and in Table 6, we see that those five factors make up the best performing model: Following Pause, Contraction Rate, Preceding Stress, COCA Following TP, and Child Age. Before shrinkage, all five of these factors were significant, after shrinkage Contraction Rate and Following Pause, the best performing factors, remain significant, as seen in Table 5.

Table 5. CDS Duration Multimodel Inferencing Results

Predictor	$\bar{\beta}$	$\bar{\sigma}$	LoCI	HiCI	p	CP
(Intercept)	0.775	0.131	0.517	1.032	0.000	NA
Following Pause	0.590	0.051	0.490	0.690	0.000	1
Contraction Rate	0.213	0.079	0.059	0.368	0.007	0.96
Preceding Stressed Syllable	-0.047	0.033	-0.112	0.018	0.157	0.80
Following COCA TP	-0.034	0.027	-0.087	0.018	0.199	0.76
Child Age	-0.001	0.001	-0.003	0.001	0.226	0.74
Following COCA JP	-0.007	0.009	-0.026	0.011	0.452	0.52

Note: reported values are coefficients with shrinkage and adjusted standard error, $n = 622$.

⁸ Forward (step-up) selection was not needed in model building as initial predictors were already determined in the random forest analyses.

Table 6. Models of CDS duration with a Δ below 2⁹

Model factors	k	df	log likelihood	AICc	Δ	w
13456	5	9	-17.96	54.22	0	0.22
123456	6	10	-17.01	54.37	0.15	0.2
12456	5	9	-18.72	55.73	1.51	0.1
3456	4	8	-19.99	56.22	2	0.08
(Null)	0	4	-97.54	203.15	148.93	0

Note: Cutoff: $\Delta < 2$, 1 = Age of Child, 2 = Following JP from COCA, 3 = Following TP from COCA, 4 = Contraction Rate, 5 = Following Pause, 6 = Preceding Stress

Positive coefficients in Table 5 are associated with longer normalized durations, and negative coefficients are associated with shorter normalized durations. A following pause (or filled pause or disfluency) results in a longer duration. The following pauses are in contexts where the speaker is searching for a word. This is consistent with research showing that when speakers have trouble accessing a word, they lengthen previous words (Bell et al., 2003; Schnadt, 2009). A higher speaker contraction rate results in longer normalized durations. This result contradicts my initial hypothesis that speakers who contract more are also likely to phonetically reduce words at a higher rate. In fact, it seems the opposite is true: speakers who contract more, phonetically reduce words at a lower rate. The adult speakers in this corpus show a strategy for compressing the time

⁹ There are a number of ways to limit the set of models to only the most predictive ones, as discussed in Burnham and Anderson (2002: 170-171). One approach is to select the top models such that the sum of their Akaike weights (probabilities) is just over 0.95. Another possibility is to use ratios of Akaike weights, where models with weights below 1/8 (0.125) of the best model are not considered. Another heuristic is to use a cutoff on ΔAIC_C values. ΔAIC_C values between 0 and 2 indicate ‘substantial’ level of empirical support for the models with those values. However, Burnham and Anderson (2002: 131) also note that a model may be within 2 units of the best model on the ΔAIC_C scale and not be a serious contender as long as it has all the parameters of the best model plus one (which punishes it by 2 units of AIC) and achieves the same data coverage, as measured by log likelihood. I present models with ΔAIC_C values of 2 or less. Models with ΔAIC_C values above ten have ‘essentially no’ empirical support (Burnham and Anderson 2002: 170).

spent on function words: either contraction or reduction, but not both. I discuss this finding further below. Target words following a stressed syllable are shorter than ones following an unstressed syllable, consistent with a trochaic bias. Words with a high following transitional probability, i.e., that are in a predictable context, are shorter than words in an unpredictable context. Following probability contexts are more important than preceding probability contexts for these grammatical items (cf. Bell et al., 2009).

The final variable that has a high cumulative probability is Child Age. Caregivers use shorter normalized durations with older children. Based on previous research, we know that parents emphasize lexical words, but not grammatical words when speaking to infants. In these cases, the lexical words would make up the greatest proportion of the utterance and grammatical words should have a quite small proportion of the utterance duration. Over time, parents emphasize lexical items less, which should result in grammatical words making a larger proportion of the utterance. However, based on the current results, it seems that by the time children are five years old, this process must have finished. Caregivers use short normalized durations of auxiliary words that only get shorter as the children get older. As we know that grammatical words have very short durations in adult speech, it seems that as children get older, caregivers speak to them more and more as they would speak to other adults. Taken together, the results of the probability measures indicate that for duration in caregiver speech, transitional probability outperforms joint probability, probability measures based on frequencies from the COCA outrank probability measures based on frequencies from the Redford Corpus, and following probability outranks preceding probability.

It is to some degree surprising that a higher contraction rate results in longer grammatical words. The results seem to show that speakers use one strategy or another to shorten their grammatical words: contraction OR shortening. Both of these strategies allow the compression of predictable, grammatical words so that more content can be included per utterance. Because this factor had such a strong effect, I examine its interaction with other speaker variables, in particular syllable rate and child age. To do this I am using functions *gam()*, *te()* and *vis.gam()* in package {mcgv} (Wood, 2014). This allows the fitting of a generalized additive model (GAM) and then the smoothing and visualization of the model predictions in a contour plot. Figure 26 shows in more detail the relationship between Speaker Contraction Rate and Speaker Syllable Rate on normalized duration. Random effects for the GAM are Speaker (Gross) and Target word. It seems clear that adults either contract more or they reduce grammatical words more, and that there is a slightly stronger effect for fast speakers ($F = 3.636, p = 0.012$), in that fast speakers who contract more have longer grammatical word durations (almost 1, meaning equal to the average syllable length of a given utterance) than slower speakers who contract more. In Figure 27, we see the relationship between Speaker Contraction Rate and Child Age on normalized duration. There is an interaction effect here ($F = 3.360, p = 0.013$). When speaking to young children, grammatical words are long no matter what the caregiver's contraction rate is, but when speaking to older children, either they use longer grammatical words and contract often or reduce their grammatical words and contraction infrequently.

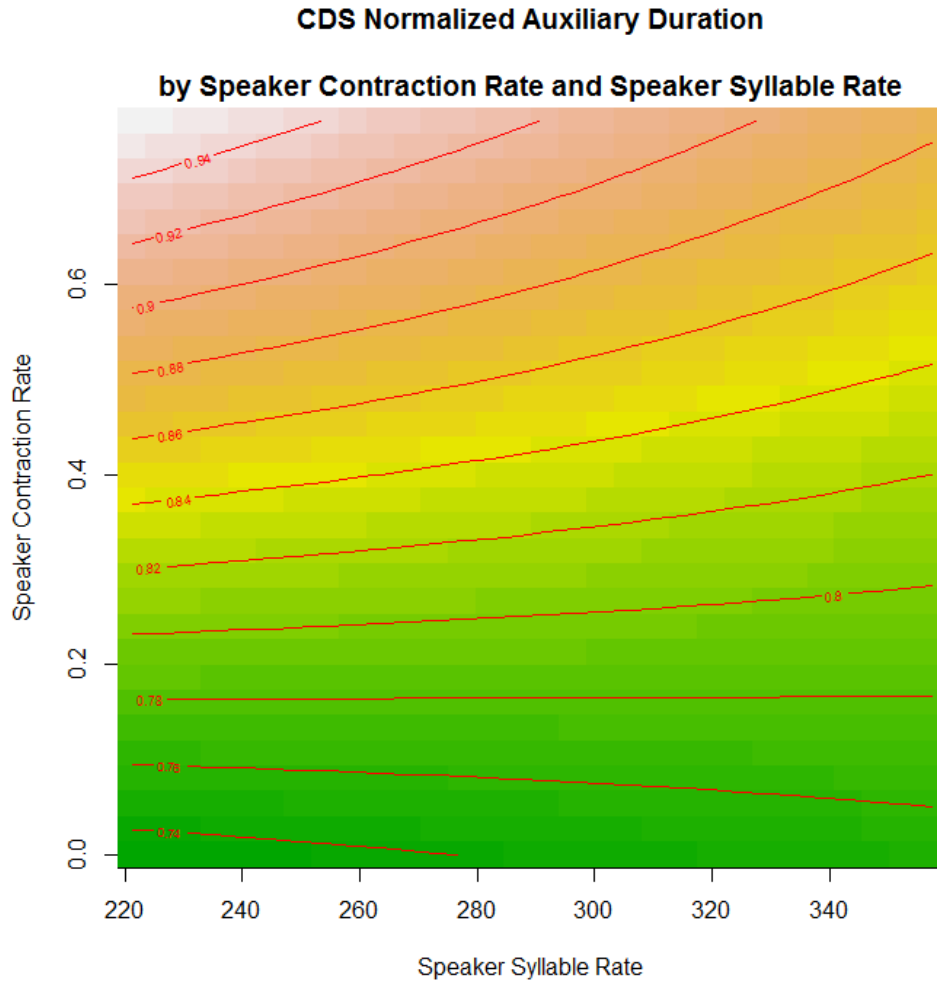


Figure 26. Interaction of Speaker Contraction Rate and Speaker Syllable Rate on CDS word duration

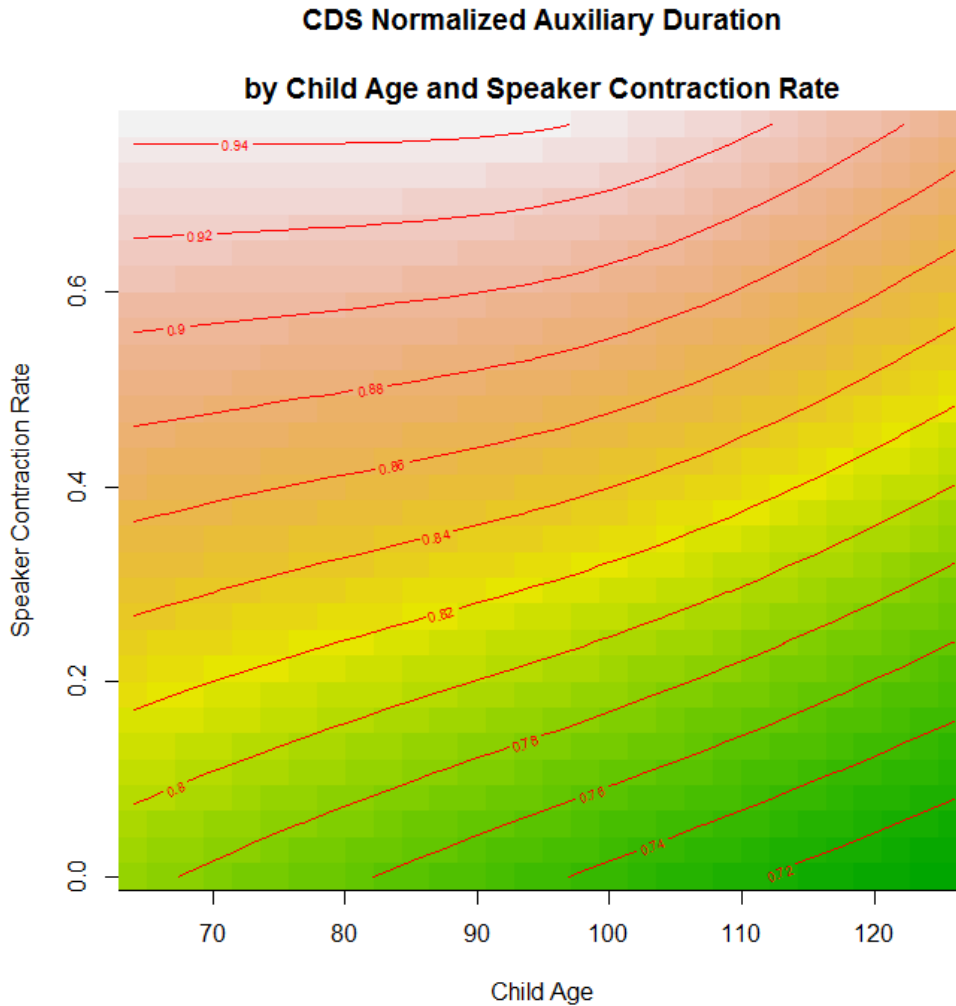


Figure 27. Interaction of Speaker Contraction Rate and Child Age on CDS word duration

Figure 28 below presents a classification tree analysis of contraction rate on word duration to determine where the best split in the data is. The partitioning of the data is at a contraction rate of 0.5. Child age was included in this analysis as a potential factor, but was not selected. This indicates a strong effect for contraction rate and a less strong effect for child age and no interaction between the two factors.

CDS Normalized Duration by Contraction Rate and Child Age

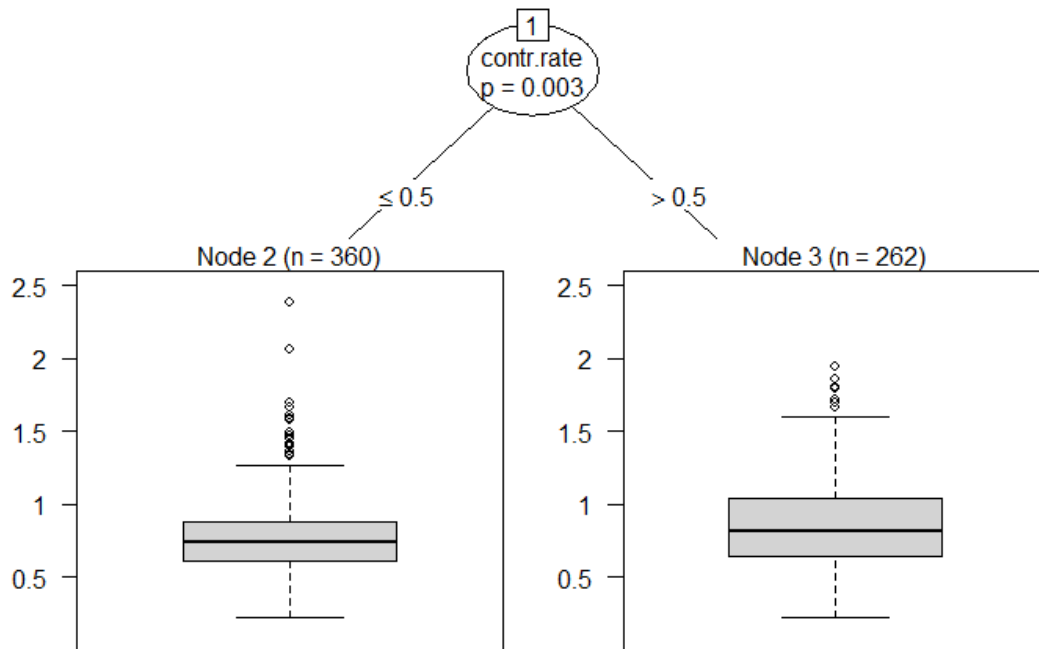


Figure 28. Partitioning of Speaker Contraction Rate on CDS word duration

No analysis for *BE* words are presented. There were no interactions with specific inflections of *BE*, factors specific to *BE* such as construction and preceding *BE* type were not significant and significant factors for *BE* word were no different than significant factors for all words combined. However, there are significant construction effects for *HAVE* words. There were no interactions with specific inflections of *HAVE*, so inflection was used as a random effect along with speaker. Factors that had a cumulative probability below 0.4 were eliminated in a backward selection procedure. As seen in Tables 7 and 8, child age, following COCA JP and contraction rate were not significant for this data set. Perfect auxiliaries are significantly shorter than modal semi-auxiliaries or possessive verbs.

Table 7. CDS *HAVE* Duration Multimodel Inferencing Results

Predictor	β^-	σ^-	LoCI	HiCI	p	Cumulative Probability
(Intercept)	0.891	0.126	0.643	1.138	0.000	NA
Following Pause	0.545	0.103	0.344	0.747	0.000	1
Construction:						0.70
Modal (reference level)						
Perfect	-0.111	0.094	-0.294	0.073	0.237	
Possessive	-0.008	0.049	-0.103	0.087	0.865	
Preceding Stressed Syllable	-0.058	0.065	-0.185	0.070	0.375	0.59
Following COCA TP	-0.031	0.046	-0.122	0.060	0.505	0.48

Note: reported values are coefficients with shrinkage and adjusted standard error, $n = 138$.

Table 8. Models of CDS *HAVE* duration with a Δ below 2.

Model factors	k	df	log likelihood	AICc	Δ	w
234	3	8	-5.1	27.31	0	0.27
23	2	7	-6.49	27.84	0.53	0.21
134	3	7	-6.72	28.31	1.00	0.17
1234	4	9	-4.79	28.99	1.68	0.12
123	3	8	-6.15	29.41	2.10	0.10
13	2	6	-8.39	29.42	2.11	0.10
(Null)	0	4	-22.72	53.74	26.43	0

Note: Cutoff: $\Delta < 2$, 1 = Following TP from COCA, 2 = Construction, 3 = Following Pause, 4 = Preceding Stress

After shrinkage, only Following Pause is significant for caregiver reduction of duration of *HAVE*. However, the most predictive model includes Construction and preceding stress as well. The model containing only the significant predictor of Following Pause is not included in Table 8, as it has a Δ of 23.16, showing that some additional predictors need to be included to achieve a predictive model. The predictors with high probability of being truly predictive are Following Pause, Construction and Preceding Stress. Adults, then, have differences in their duration based on the meaning of *HAVE*, as well as the presence of a following pause. There are only 8 instances where *HAVE* is followed by a pause, but these instances are significantly longer than the others. All of these instances occurred when a speaker was searching for a word. The longer durations then, are likely used by speakers to buy time to access the word they are trying to find (Bell et al., 2003; Kapatsinski 2010). Caregivers produce shorter durations when *HAVE* is a perfect auxiliary than when *HAVE* is the modal semi-auxiliary. The durations of possessive verbs are also significantly longer than the durations of perfect auxiliaries, but the modal semi-auxiliary and the possessive verbs are not different from one another. The reduced auxiliary duration distinction shows a sensitivity to grammatical status, as the lexical possessive verb is longer than the grammatical perfect auxiliary, even though the lexical word is more frequent than the grammatical word. As in the model for all words, having a preceding stressed syllable results in a shorter target word, consistent with English's trochaic bias.

3.4.4. Conclusion for Word Shortening in Child-Directed Speech

Now it is possible to return with answers to the research questions presented at the beginning of this section.

Q1. Are the function words *HAVE* and *BE* subject to probabilistic reduction?

Yes, just like content words, these function words are more reduced (here operationalized as shorter in milliseconds, normalized by speaking rate of the given utterance) when they are in more probable contexts. Even for very short, frequent words, speakers are sensitive to context probability and modify their behavior accordingly. This effect could either be due to faster access to words in a more probable context, reducing the time needed for planning, resulting in faster production, or due to the speaker's expectation that the listener will understand a lenited or fast production because the listener also knows the context is probable.

Q2. Do caregivers adapt their speech in one or more phonetic characteristics to school-aged children?

Yes, caregivers not only adapt their speech to school-aged children in content, but also in the phonetic characteristics of function words. Function words are produced faster (in raw duration), as well as proportionally faster as compared to content words, as a caregiver's child gets older. Because caregivers do not adapt their speech rate to child age (cf. Figure 15), but do adapt their function word production to child age, I conclude that the caregivers are lessening their listener-accommodation to their children by not needing to spend time on unimportant function words. Because there is a change in duration over time, while still being probabilistic for all age groups, it seems that a listener oriented account of function word shortening is more appropriate than a speaker-internal account, at least for this speaker group. The effect of predictability on word accessibility within the production system of a caregiver does not change over time, as their child ages. That is, it should be just as easy for a speaker to access word from the mental lexicon in a high

probability context when speaking with a young child as with an older child, making a speaker-internal account for increased function word shortening with older children unconvincing. Note, however, that this result indicates some kind of listener sensitivity on the part of the caregivers, not necessarily one implemented as online monitoring and modeling of the addressee.

Q3. Do adults who contract more often also shorten grammatical words to a greater degree than those who contract less often?

No, adults use one reduction strategy or another: they either shorten words (and do not contract them) or they contract words (and do not shorten them). Both of these strategies are means to compress unimportant, predictable information into a short time window, so that more time can be spent on important, unpredictable information. Adults choose one or the other of the strategies.

Q4. Do adults differentiate function word production by construction (meaning)?

Sometimes. Productions of *BE* do not differentiate in word length in relation to their meaning. However, productions of *HAVE* do differ in word length in relation to meaning. Perfect auxiliaries are significantly shorter than other meanings of *HAVE*, even when other variables are controlled for.

3.5. Word Shortening in Child Speech

3.5.1. Introduction

The analyses of word shortening in child speech are aimed to answer the following primary research questions:

Q5 Are the function words *HAVE* and *BE* in child speech subject to probabilistic reduction?

Q6. Does school-aged children's function word production change over time?

Q7. Do children who contract more often also shorten grammatical words to a greater degree than those who contract less often?

Q8. Do children differentiate function word production by construction (meaning)?

Q9. Are the factors that influence word shortening in child speech different than those factors that influence word shortening in their input (caregiver speech)?

3.5.2. Random Forest Variable Importance Rankings for Word Shortening in Child Speech

Figures 29 through 34 present the variable importance ranking for children's speech. These figures show that the relative importance of preceding context and speaker-based variables stay fairly stable over time, but that there is quite a bit of variation in the ranking of following context variables. There are also several variables that are influential for children but not for caregivers, particularly: a preceding pause influences duration for children but not caregivers and a slower syllable rate, associated with less developed motor control for children, also increases duration for children but not caregivers.

Figure 29 shows the importance of speaker- and word-specific variables. As with caregivers, the particular lemmas and word forms make a difference in normalized duration. Although they are all monosyllabic words, they differ in their phonemic structure, leading to differences in duration relative to the rest of the syllables in the utterance in which they occur. Utterance Quartile is also ranked as important, just as it was in the importance ranking for caregivers, although it was not significant in the regression models for caregiver duration. Speaker is also ranked as an important variable,

but Speaker (Gross) and Speaker (by Text) are ranked at relatively the same importance. As Figure 23 showed, for adults, Speaker (Gross) was ranked much higher than Speaker (by Text), indicating that there is more stability within the behavior of a particular speaker over time in adult productions, and more variability within the behavior of a particular speaker over time in child productions. As with adult speech, Contraction Rate is an important variable in child speech. Unlike for caregivers, Speech Rate is an influential variable for normalized duration of grammatical words. For children, this variable is associated with developmental ability of motor control and correlates with age. Speaker Age (in Months) and Grade also reached significance, but were ranked lower than Speech Rate, showing that actual development of motor control ability is more important than simple age or grade, which are heuristics for developmental ability, for predicting reduced duration.

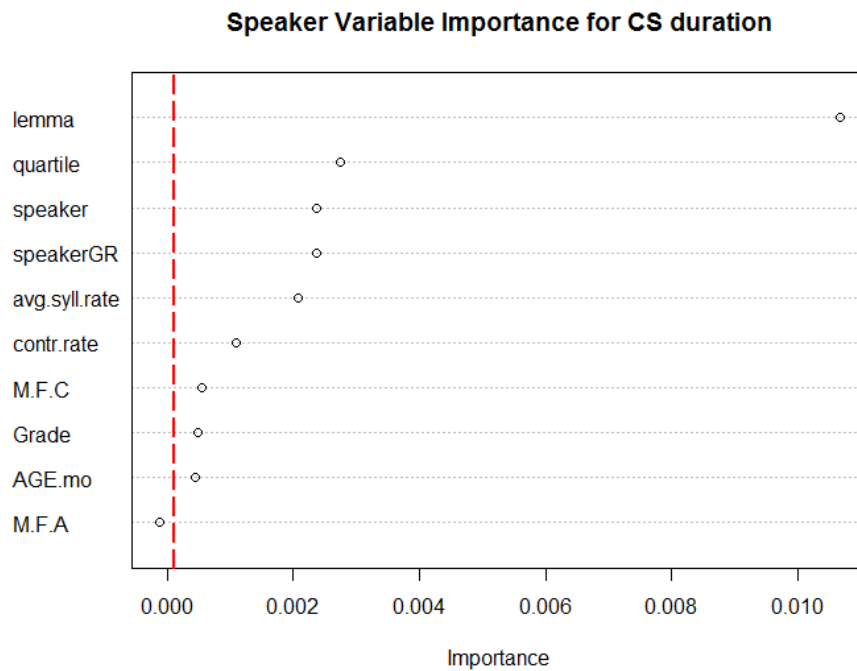


Figure 29. Speaker-specific and word-specific variable importance for CS grammatical word duration

Figure 30 shows the development of variable importance rankings over time for the speaker-specific and word-specific factors. For every grade group, the phonemic structure of the word is important, just as it is for adults. Longer words take longer to say, although there is variability, as seen in Table 2, the shortest and longest words in duration are not necessarily the shortest and longest words in phonemic structure. Only the kindergarten-1st grade and the 4th grade-5th grade groups had data points from the same speaker from different texts, so only for these two grade groups were both Speaker (Gross) and Speaker (by Text) included as factors to rank. For the kindergartners and 1st graders, neither is important. For the 4th and 5th graders, Speaker (Gross) outranks Speaker (by Text) in importance, mirroring the importance ranking found in adult speech. Age (in months) is important only for the youngest group. This also shows that there is some stability in reduction behavior for the older group, as there is not as much variance due to specific ages. Utterance quartile is ranked as important for the youngest three groups. However, all factors ranked as important are only barely important in comparison to the importance of the word form. This indicates we should not expect to see a great deal of significance for speaker- or word-specific factors, and where we find significance, we should not expect to see a strong interaction with age of the child.

Figure 31 shows the importance ranking for preceding context variables on normalized word duration. Preceding Transitional Probability based on frequencies from the COCA is ranked as the most important variable. For caregivers, as shown in Figure 24, it was Preceding Joint Probability. For preceding elements, transitional probability here is a “forward transitional probability”: how likely is it that *is*, *was*, *have*, etc. is going to come next given that the preceding word is *he*, *there*, *we*, etc. A high preceding

transitional probability example from this dataset is *there was* (-0.923). Given *there, was* is a likely word to come next. A low preceding transitional probability example from this dataset is *turtle were* (-3.465). Given *turtle, were* is not a particularly likely word to come next. When preceding transitional probability is calculated based on frequencies from the Redford Corpus, bigrams like *there was* still have high transitional probability (-0.495), but bigrams like *turtle were* have higher transitional probabilities than in the COCA (-1.708) as proportionally, there are more instances of turtles doing things in the Redford Corpus than in the COCA. Preceding transitional probability based on Redford Corpus frequencies is also ranked as important, but less than the transitional probability based on frequencies from the COCA. Whereas both Joint Probability variables were ranked as influencing grammatical word duration for caregivers, only the Joint Probability based on Redford Corpus frequencies is ranked as important for influencing duration for children. The more probable the bigram, the shorter the target tends to be. The range of probabilities here for preceding joint probability based on Redford Corpus frequencies is small (-2.402 to -4.892). The three most highly probable bigrams from this data subset are *he was* (-2.402), *he is* (-2.487) and *frog was* (-2.569). Less probable bigrams occur only once and include errors such as *a was*, *was be*, and *goes has* (-4.892).

Figure 31 also shows that the priming variable Preceding *BE* Type significantly influenced duration, although Preceding *HAVE* Type and Preceding Contraction did not, just as it was for adults. Preceding *HAVE* Type probably has less influence because there are few tokens of *HAVE* in the corpus to actually precede the targets. As with adult speakers, when the word preceding the target is stressed, the target is somewhat shorter in normalized duration than when the preceding word is unstressed. Children, like adults,

are motivated by stress patterns, with an unstressed function word likely to be even shorter when appearing after a stressed syllable. Finally, a preceding pause results in a longer duration for targets in children’s speech, although it does not in caregiver speech. Children either have long utterance-initial words in general, or when an utterance-initial word is a grammatical word which normally is utterance-medial, it could be a sign they are still planning and are buying time with longer durations to allow for further time to access the upcoming word.

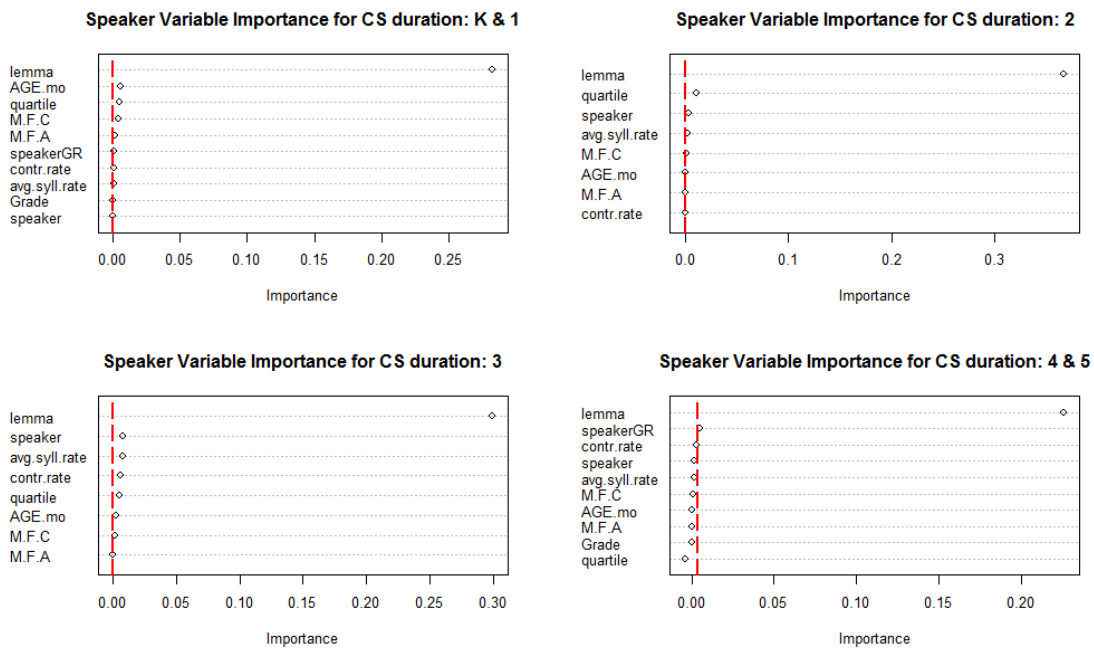


Figure 30. Speaker-specific and word-specific variable importance for CS grammatical word duration by school grade

Preceding Context Variable Importance for CS duration

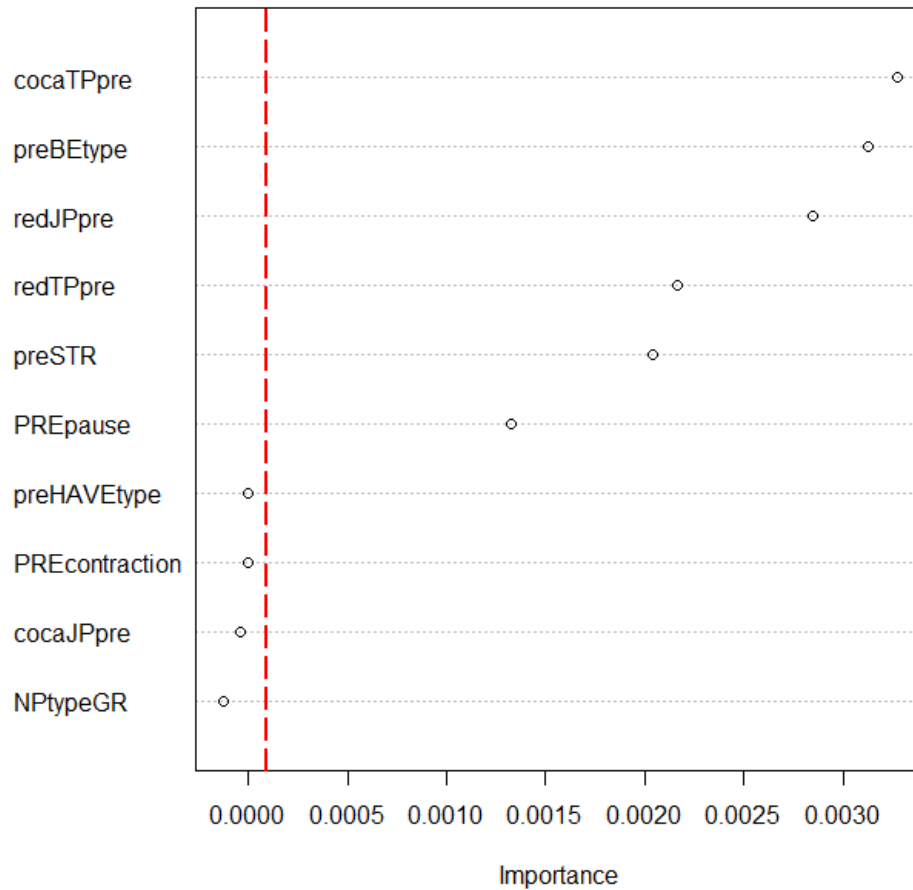


Figure 31. Preceding Context variable importance for CS grammatical word duration

Figure 32 shows the development of variable importance rankings over time for preceding context factors. The ranking of preceding context variables are more stable over time than the ranking of speaker-specific variables. For all grade groups, the two highest ranked variables, ranked at relatively the same position, are 1) Preceding Transitional Probability based on frequencies from the COCA and 2) Preceding *BE* type, which is a priming variable. For all grade groups the next best performing variable is Preceding Joint Probability based on frequencies from the Redford Corpus, followed by

Preceding Transitional Probability based on frequencies from the Redford Corpus, a Preceding Stress Context and finally Presence of a Preceding Pause. A preceding pause was not a significant factor influencing duration for adult speakers.

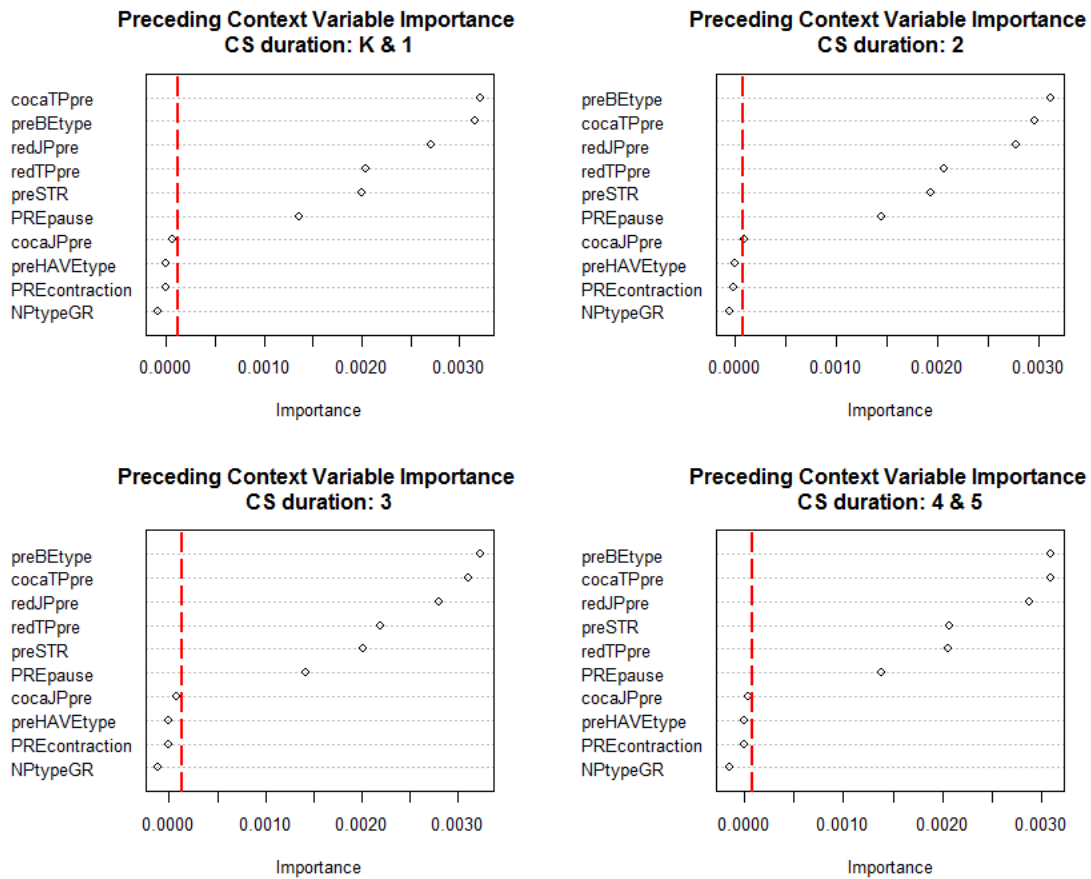


Figure 32. Preceding Context variable importance for CS grammatical word duration by school grade

Figure 33 shows the variable importance ranking for following context variables on children’s normalized word durations. Results from children’s productions are similar

to results from adults' productions. The most important variable is the presence of a following pause, indicating the effects of the need for planning time to access an upcoming word and final lengthening. Following Transitional Probability based on COCA frequencies is rated next most highly. Following transitional probability here is a "backward transitional probability": how likely is it that the preceding word is *have*, *is*, *was*, etc. if the following word is *got*, *still*, *glad*, etc. A high following transitional probability example from this dataset is *have got* (-0.655). Given the occurrence of *got*, the preceding word is likely to be *have*. A low following transitional probability example from this dataset is *has water* (-3.626). Given the occurrence of *water*, the preceding word is unlikely to be *has*. Construction meaning is also an influential variable on duration. As with caregivers, Transitional Probability based on Redford Corpus frequencies and Joint Probability based on COCA frequencies are also ranked as important, although in the reverse order: Redford Corpus Following Transitional Probability was ranked as more important for child speech. Joint Probability based on Redford Corpus frequencies does not rank as important for children's productions. A following stressed syllable means that the target is somewhat likely to be shorter, but this is not ranked much higher than the importance threshold.

Figure 34 shows the longitudinal development for following context variable importance in children's speech. The most important variable for both child and caregiver speech overall is the presence of a following pause. This variable is ranked as important only for 2nd and 3rd graders. For all of the groups, except for third graders, the next most important variables paralleled the overall analysis, as Construction Meaning and Following Transitional Probability based on either frequencies from the Redford Corpus

or COCA, are all highly ranked, however their relative ranking differs. Following Joint Probability based on frequencies from the Redford Corpus is ranked as important for all of the groups except for the oldest. A following stress context is important only for the oldest two groups. Joint Probability based on frequencies from the COCA is ranked as important, but barely, for all of the groups except for third graders.

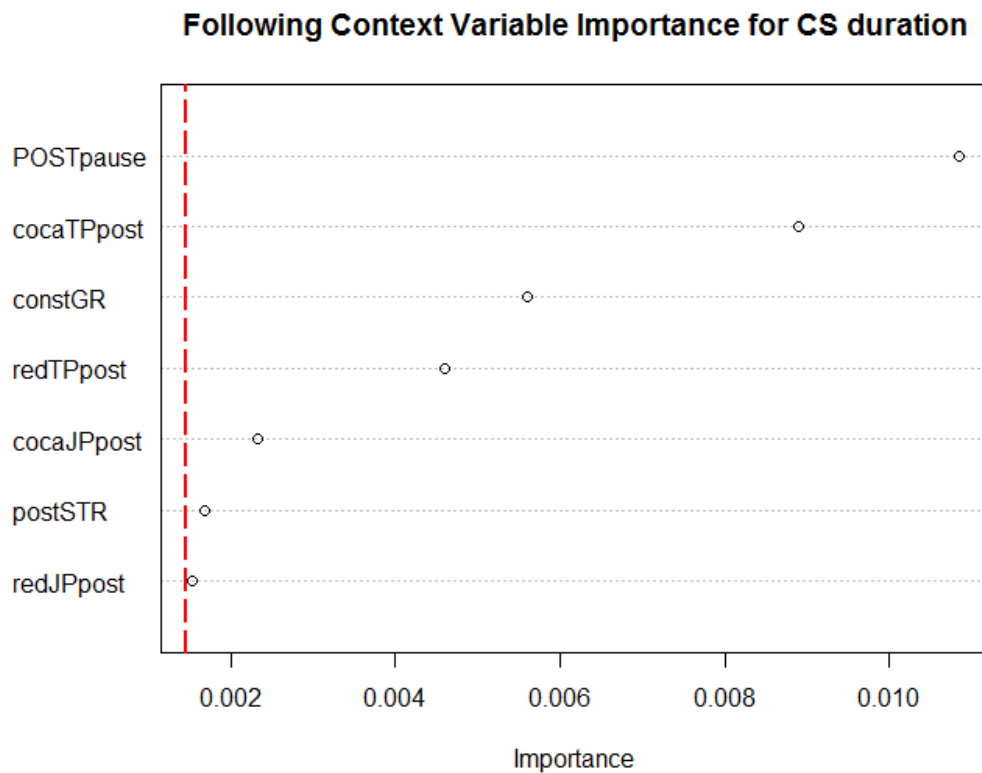


Figure 33. Following Context variable importance for CS grammatical word duration

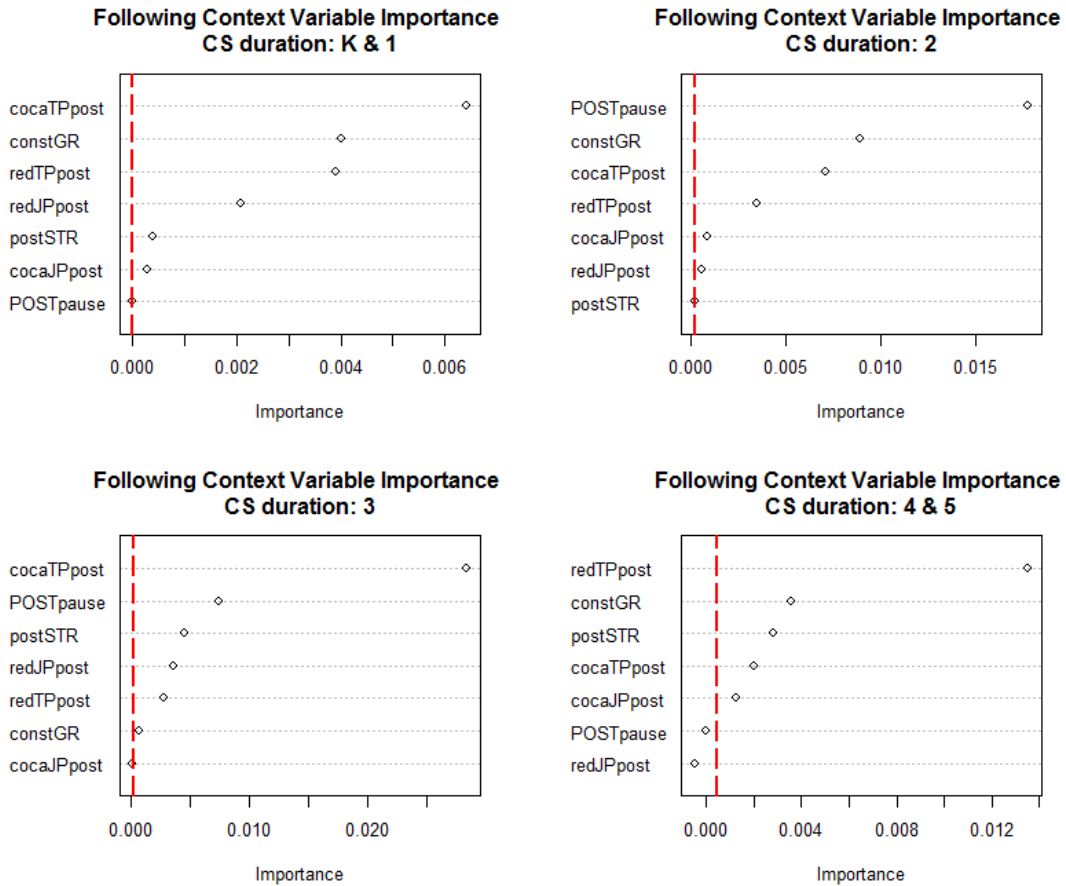


Figure 34. Following Context variable importance for CS grammatical word duration by school grade

Although a Following Pause is predictive of longer durations in Caregiver Speech (cf. Figure 25) and in the speech of children in other grades, it is not predictive for the youngest grade group, as there are very few instances of a target grammatical word being following by a pause (3 followed by a pause v. 174 not followed by a pause).

For the first three grade groups, transitional probability from COCA frequencies outranks transitional probability from Redford Corpus frequencies, but for the oldest grade group, the relationship is reversed.

3.5.3. Multimodel Inferencing Analyses for Word Shortening in Child Speech

Models for child speech included an interaction term for age, in order to see the developmental trajectory of reduction and the developmental trajectory of factors affecting reduction. All possible non-collinear variables were included in models, with an interaction term for age and then backwards selection procedures were used until only factors (or interactions) with a cumulative probability over 0.4 remained. First all words were combined in one analysis, checking for interactions with specific words or inflections. None were found and so Lemma was used as a random effect. Second, analyses were conducted for *BE* and *HAVE* words, testing for effects of priming and construction. None were found for *BE*, but paralleling the results in caregiver data, there were construction specific effects for *HAVE*, therefore results for *HAVE* are also presented below. After presenting the results from the models, I discuss the differences between the results for caregivers and their children.

The first model built to compare normalized word durations for children combines all word types: *are* ($n = 18$), *be* ($n = 14$), *had* ($n = 67$), *has* ($n = 16$), *have* ($n = 9$), *is* ($n = 97$) *was* ($n = 420$) and *were* ($n = 100$), (total $n = 741$). Random effects for this model are Speaker (Gross) and Lemma. Factors that were ranked as important in the random forest analyses were Preceding and Following Transitional Probability based on COCA frequencies, Utterance Quartile, Preceding and Following Pause, Average Syllable Rate, Speaker Contraction Rate and Speaker Gender. Eliminated variables that were ranked as important in the random forest analyses, but were not ranked as probable in multimodel averaging include: Speaker Gender, Speaker Syllable Rate, Preceding

Pause, COCA preceding Transitional Probability, Preceding Stress, COCA following Joint Probability.

As seen in Table 9 below, a probable following context results in a shorter target word and a following pause or disfluency results in a longer target word. Both of these variables have a high cumulative probability (1), indicating that these variables always have a reducing and lengthening effect, respectively, no matter the age of the child. A following pause results in a longer duration, as was the case for caregivers. Second graders, as well as third, fourth and fifth graders, have proportionally more instances of a pause following a target grammatical word than the adults (2nd graders: 6.77%, 3rd graders: 7.42%, 4th and 5th graders: 6.50%, caregivers: 4.03%), indicating that they may be having more difficulty searching for upcoming words than the adults, which is to be expected. The youngest grade group has a lower percentage of disfluencies following a target, but it should be noted that they are also speaking considerably slower than the other groups, meaning they have more time to plan without stopping and restarting an utterance.

A speaker with a high contraction rate has shorter normalized durations, as the coefficient is negative. The effect of speaker contraction rate on normalized duration was positive for the caregivers. The interaction term of speaker contraction rate and speaker age is also highly probable and is positive for children. This indicates that the effect of contraction rate on reduction is attenuated over time. The direction and strength of this interaction is investigated more fully below (cf. Figure 36 below).

Finally, we see that as children get older, their normalized auxiliary durations also are shorter. Proportionally, less time is spent on the grammatical words in the utterances

in which they occur. However, because duration is also influenced by context probability for children, we know that it is not only better speech production, but also better sensitivity to meaning and context that results in shorter word durations. Finally, words that are in the final quartile of an utterance are longer than words in any of the other quartiles. There is also a negative interaction between Utterance Quartile and Age, indicating that as children get older, the lengthening effect in the final quartile is diminished. This indicates that younger children are more likely to engage in word lengthening in the final utterance quartile than older children. There was no effect of lengthening in the final utterance quartile for caregivers, indicating that this effect will likely continue to diminish over time for function words.

The model ranking output (Table 10) shows that there is low model selection uncertainty for the given predictors. The weights of the top three models add up to 0.93, all other remaining possibilities have a high Δ and very low w . The best performing model includes all of the predictors. The next best model, which lacks the main effect and interaction for Utterance Quartile, fares worse. Therefore, we can be confident about the predictiveness of the best performing model.

Figures 35 and 36 show in more detail the relationship with age and contraction rate and syllable rate and contraction rate for normalized duration. In Figure 35, we see that children who speak quickly have chosen a reduction strategy: they either contract often, but do not reduce grammatical words or they reduce grammatical words, but do not contract them ($F = 3.642, p = 0.012$). Slow speaking children either contract and reduce word duration or do neither. Syllable rate and age are tied together for children, but we see that there is a stronger effect for age and contraction rate ($F = 5.966, p < 0.000$) than

there is for syllable rate and contraction rate on normalized word duration. Figure 36 shows that the effect for age is quite clear, older children either contract or reduce, younger children either contract and reduce or neither. The pattern for the older children, then, matches the pattern of adult usage. Sometime around age 7;6 (90 months), children choose a strategy for shortening predictable, grammatical words: contract or reduce.

Table 9. Child Speech Duration Multimodel Inferencing Results

Predictor	β	σ	LoCI	HiCI	p	CP
(Intercept)	0.882	0.263	0.366	1.397	0.001	NA
COCA Following TP	-0.083	0.021	-0.125	-0.042	0.000	1
Following Pause	0.325	0.053	0.222	0.428	0.000	1
Speaker Contraction Rate	-0.851	0.438	-1.710	0.008	0.052	0.99
Speaker Age	-0.003	0.003	-0.008	0.002	0.259	0.97
Age by Contraction Rate	0.010	0.004	0.001	0.019	0.026	0.93
Quartile 1 (Reference Level)						
Quartile 2	0.059	0.159	-0.254	0.371	0.712	0.66
Quartile 3	0.063	0.164	-0.257	0.384	0.698	
Quartile 4	0.331	0.388	-0.429	1.091	0.394	
Age by Quartile 1 (Reference Level)						
Age by Quartile 2	-0.001	0.002	-0.004	0.003	0.687	0.43
Age by Quartile 3	-0.001	0.002	-0.004	0.003	0.725	
Age by Quartile 4	-0.003	0.004	-0.010	0.004	0.440	

Note: reported values are coefficients with shrinkage and adjusted standard error, $n = 741$.

Table 10. Models of Child Speech duration with a Δ below 2.

Model factors	k	df	log likelihood	AICc	Δ	w
1234567	7	15	-82.4	195.46	0	0.41
12346	5	9	-88.88	196.01	0.55	0.31
123456	6	12	-86.16	196.74	1.29	0.21
(Null)	0	4	-136.25	280.56	85.1	0

Note: Cutoff: $\Delta < 4$, 1 = Speaker Age, 2 = Following COCA TP, 3 = Speaker Contraction Rate, 4 = Following Pause, 5 = Utterance Quartile, 6 = Speaker Age by Contraction Rate, 7 = Speaker Age by Utterance Quartile

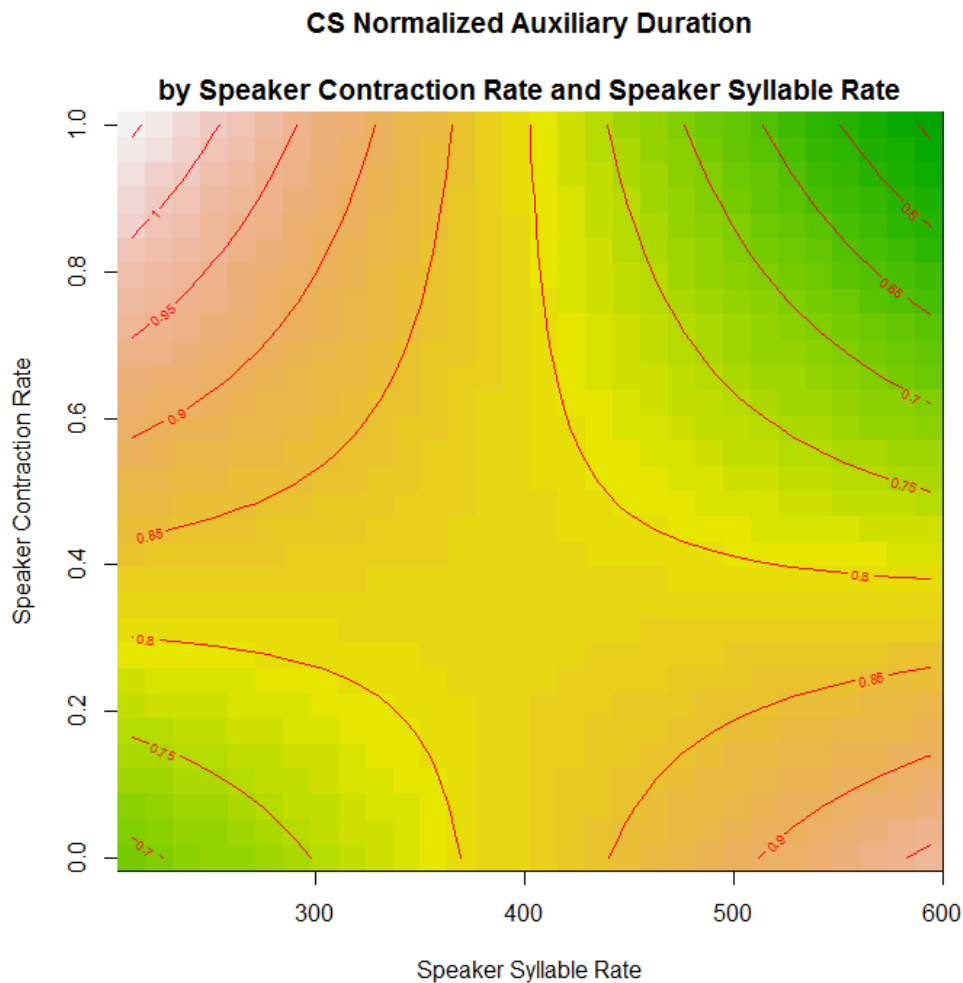


Figure 35. Interaction of Speaker Contraction Rate and Speaker Syllable Rate on CS word duration

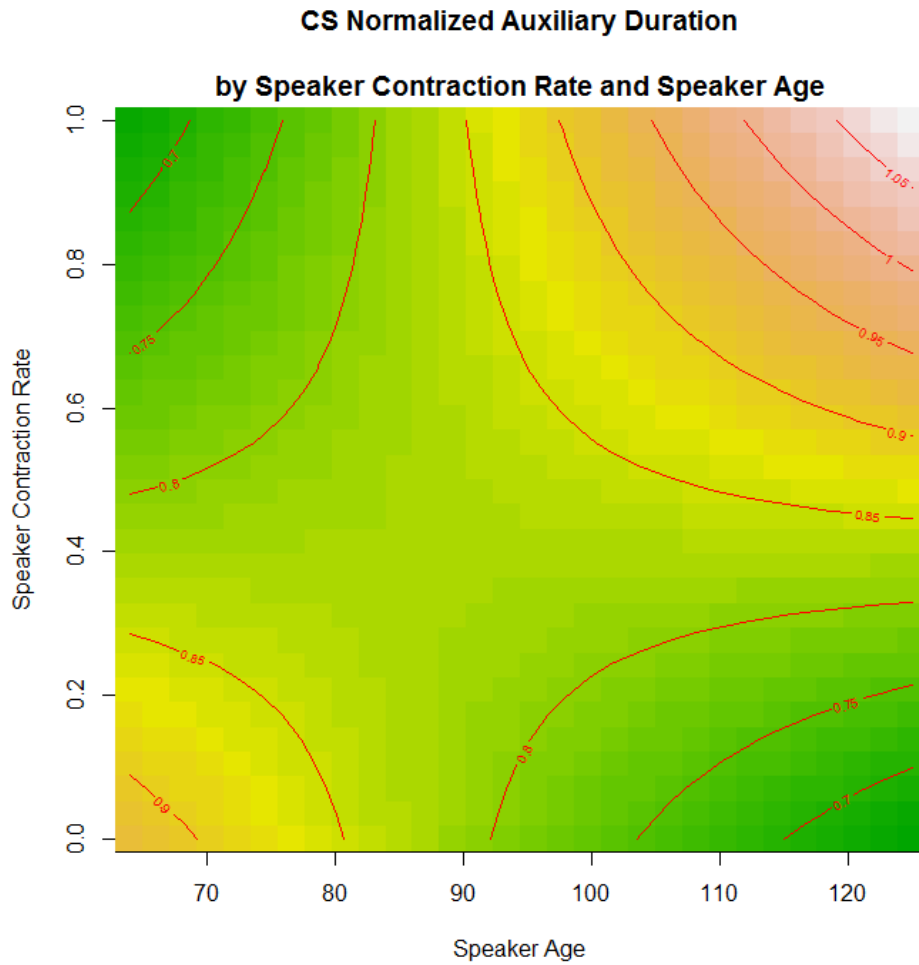


Figure 36. Interaction of Speaker Contraction Rate and Speaker Age on CS word duration

Figure 37 examines the relationship between parent and child contraction rates. For a subset of children and caregivers there are dyads, we have texts from both child and their caregiver (65 pairs of texts). For these dyads, we can see that there is no correlation between caregivers' contraction rate and those of their children. This indicates that children choose a reduction strategy on their own that is independent of the reduction strategy of their parents.

Child-Caregiver Dyad Contraction Rate Correlation

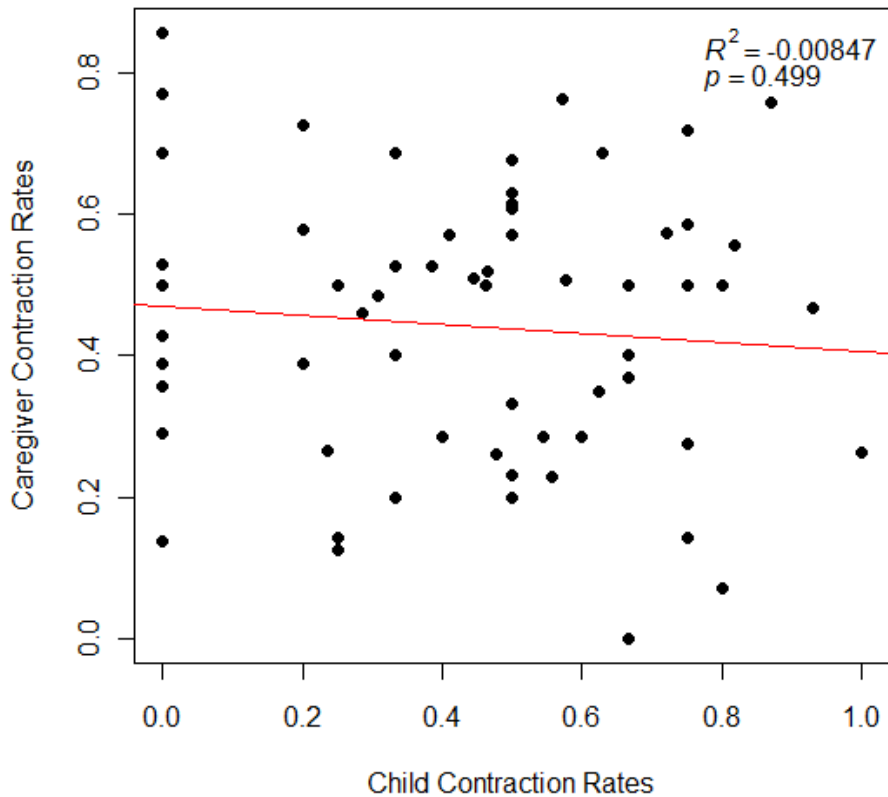


Figure 37. Correlation of Caregiver and Child Dyad Contraction Rates

In summary, for both children and caregivers, the age of the child matters, with older children hearing and producing proportionally shorter function words. For both caregivers and children, a following probable context results in a shorter word. Contraction Rate is important for both caregivers and children, but with child behavior becoming more like caregiver (adult) behavior as the child gets older. Quartile is important for children, especially for young children with its importance diminishing over time. These results show that as children get older, their behavior becomes more adult-like in function word production.

Only a few factors were significant for *HAVE*. There was a small n here, only 98 cases, so when there are significant effects it is possible to have confidence the effects are strong. As seen in Table 11, there are no interactions for age for this data set, also possibly due to the low n . There are main effects for Following Transitional Probability, with more following contexts resulting in shorter words. There is a main effect for Construction, with perfect auxiliaries being significantly shorter than modal semi-auxiliaries or possessive verbs, just as there was for caregivers. Preceding stress is significant, with targets following a stressed syllable being shorter than targets following an unstressed syllable. Finally, targets occurring in the final quartile of the utterance are significantly longer than targets occurring in other quartiles. As seen in Table 12 further below, there is fairly low model selection uncertainty, with the best performing model including all factors and having a w of 0.49.

Table 11. CS *HAVE* Duration Multimodel Inferencing Results

Predictor	β	σ	LoCI	HiCI	p	CP
(Intercept)	0.961	0.169	0.630	1.293	0.000	NA
COCA Following TP	-0.064	0.056	-0.174	0.046	0.252	0.70
Construction						0.99
Modal (Reference Level)						
Perfect	-0.340	0.103	-0.542	-0.139	0.001	
Possessive	-0.053	0.079	-0.207	0.101	0.497	
Preceding Stressed Syllable	-0.136	0.098	-0.329	0.056	0.164	0.79
Quartile						0.89
Quartile 1 (Reference Level)						
Quartile 2	-0.028	0.074	-0.172	0.117	0.709	
Quartile 3	0.077	0.075	-0.071	0.224	0.309	
Quartile 4	0.242	0.122	0.002	0.482	0.048	

Note: reported values are coefficients with shrinkage and adjusted standard error, $n = 98$.

Table 12. Models of CS *HAVE* duration with a Δ below 2.

Model factors	k	df	log likelihood	AICc	Δ	w
1234	4	11	-4.83	35.05	0	0.49
234	3	10	-6.99	36.76	1.71	0.21
(Null)	0	4	-22.66	53.8	18.75	0

Note: Cutoff: $\Delta < 2$, 1 = Following TP from COCA, 2 = Preceding Stress, 3 = Construction Meaning, 4= Utterance Quartile

The results for *HAVE* dataset in child speech parallel the results for child-directed speech. Utterance Quartile is only significant for children (as with the combined words analysis) and Following Pause is only significant for caregivers, but there were only five instances of *HAVE* followed by a pause, so there may not have been enough cases for significance to be reached, as it was in the combined word analyses for both groups.

3.5.4. Conclusion for Word Shortening in Child Speech

It is now possible to return with answers to the research questions presented at the beginning of this section.

Q5. Are the function words *HAVE* and *BE* in child speech subject to probabilistic reduction?

Yes, even the short and highly frequent words *BE* and *HAVE* are subject to probabilistic reduction in child speech.

Q6. Does school-aged children’s function word production differ over time?

Yes. Function words get shorter over time, as is clear through examining raw measures of duration, and over time function words get proportionally shorter compared to content words. Increased syllable rate does not account completely for the raw reduction in milliseconds needed to produce these words. Even though duration measures

are normalized for syllable rate, older children still produce shorter function words than younger children. Older children have learned that these kinds of predictable words are still understood by their listeners and less effort can be expended in their production. If the shorter function words were due only to increased motor skill in producing unstressed syllables, then children would not produce variable durations according to context probability. Because their productions do reflect a sensitivity to context probability, that is a sign that word predictability, not just motor skills, affects function word production in child speech. In addition, the change in the influence of predictability over age indicates that in these five years, children become more adult-like in their function word production.

Q7. Do children who contract more often also shorten grammatical words to a greater degree than those who contract less often?

Yes and no, as it interacts with child age. Young children who reduce more also contract more. These children also have higher text entropy scores than young children who do not reduce or contract (cf. Figure 38 below for effect of age and text entropy on word shortening). This indicates that young children who use information compression early on have a better understanding of what is important and unimportant in an utterance. They are more sensitive to context. However the strength of this entropy-age interaction effect on normalized duration diminishes over time. Older children use one strategy or another for information compression, just like their caregivers, and then there is no longer a strong correlation between what kind of information compression strategy a speaker uses and their language proficiency (as inferred by text entropy).

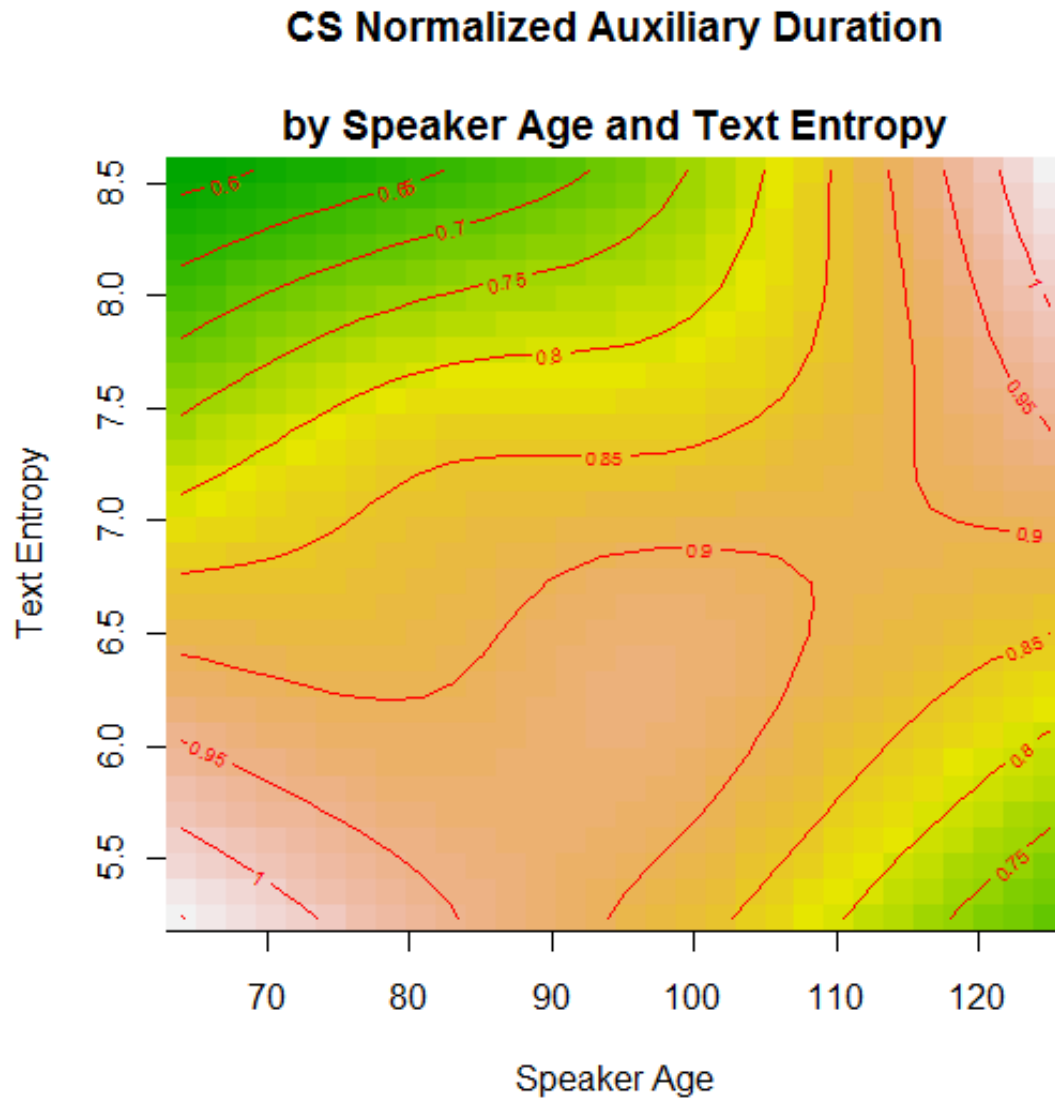


Figure 38. Child Normalized Duration by Text Entropy and Age

Q8. Do children differentiate function word production by construction (meaning)?

Sometimes. As with caregivers, *BE* does not vary in duration by construction but *HAVE* does. *HAVE* tokens in perfect constructions are significantly shorter than *HAVE* tokens in possessive or modal constructions.

Q9. Are the factors that influence word shortening in child speech different than those factors that influence word shortening in their input (caregiver speech)?

Not really. Children were not sensitive to stress context or joint probability, as adults were, and children were more sensitive to the position of a word within an utterance (Utterance Quartile), but otherwise children and caregivers were sensitive to the same factors. Where there were interactions with age, older children behaved more like adults than young children, showing their behavior becoming more adult-like.

3.6. Contraction in Child-Directed Speech

The previous four dependent variables were based on phonetic data. Contraction can be extracted from orthographic transcripts much more easily. Therefore, as described in 3.2, more data was used for the contraction analyses than for the phonetic analyses.

The data in this section is based on 74 narrations from caregivers as opposed to 44 narrations used in the analyses above. The main research questions for this section are:

Q10. Are there differences in caregiver contraction depending on the age of the child?

Q11. Are there different motivations for contraction vs. word shortening (duration reduction)?

3.6.1. CDS Contraction of *HAVE*

As seen in Figure 39, almost half of the time that caregivers produce perfect auxiliaries, they contract them. Some adults always contract, some always use full forms, but about a third sometimes contract and sometimes do not. As there are so few perfect auxiliary contractions, no statistical analyses of *HAVE* contractions were done

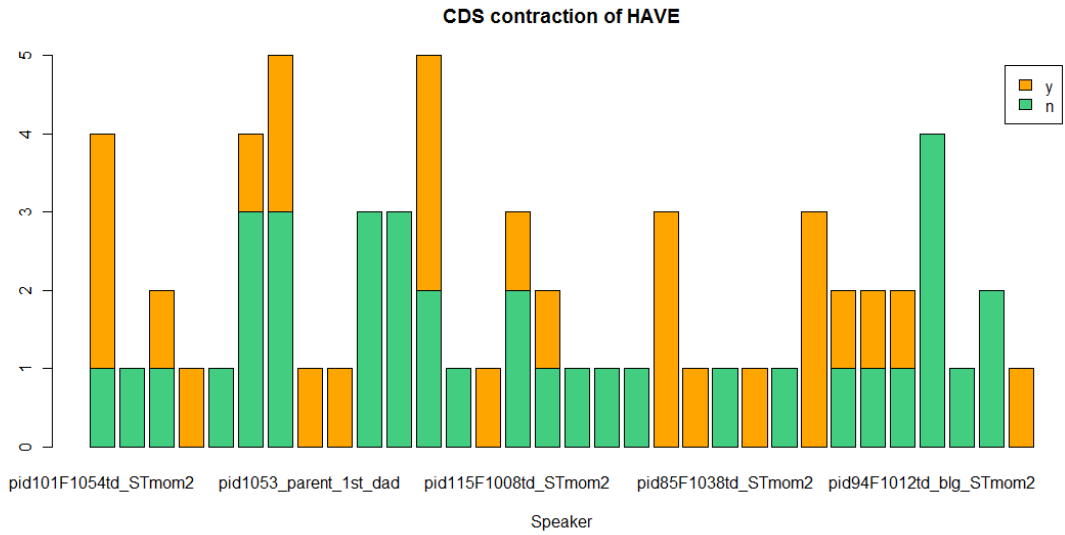


Figure 39. CDS Contraction rate of Perfect *HAVE* by Speaker

3.6.2. Contraction of *BE*

3.6.2.1. Introduction

As seen in Figure 40, most caregivers mix full and contracted forms of *BE*, although some contract at higher rates than others.

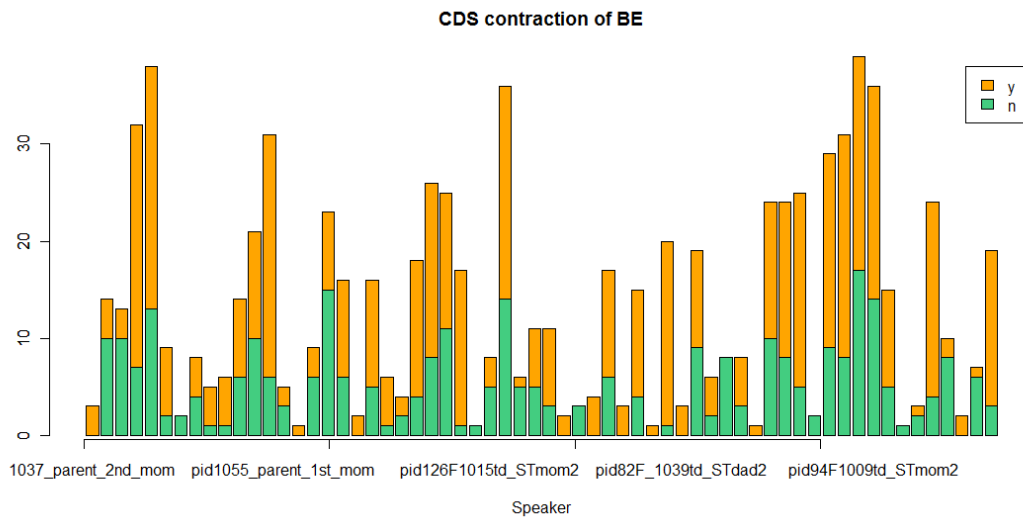


Figure 40. CDS Contraction rate of Perfect *BE* by Speaker

3.6.2.2. Random Forest Variable Importance Rankings

Here I present the results of contraction random forest analyses variable importance rankings for caregiver speech. Figure 41 shows that there is variation in contraction due to which utterance quartile the target is in, the particular inflection of *BE*, and that Speaker Contraction Rate is very important for predicting contraction, as is reasonable. Contractors contract more, non-contractors contract less. Speaker (Gross) is ranked higher than Speaker (by text), indicating that the differences between particular speakers are more important than the differences between speakers, year-to-year. Finally, we see barely significant effects for child age and child gender.

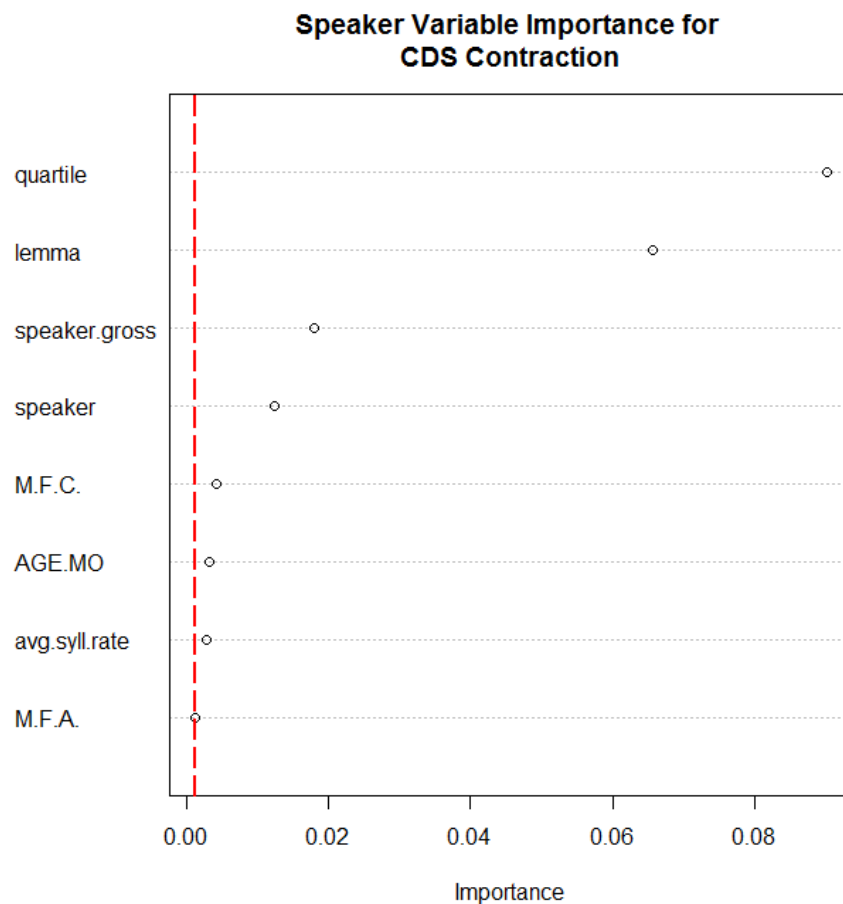


Figure 41. CDS Word and Speaker-specific factor variable importance ranking for contraction

From the random forest analyses of preceding contexts in Figure 42, it is clear that the COCA based probabilities perform much better than any of the other factors. Transitional probability performs better than joint probability, even though we may expect that contracted elements are highly associated with their preceding element leading to stored chunks. If this were the case, then joint probability should perform better than transitional probability in predicting contraction. But note that the COCA JP and TP are very highly correlated ($r = 0.84$), so it is likely that either would have a high cumulative probability in a multimodel inferencing analysis.

Preceding Context Variable Importance for CDS Contraction

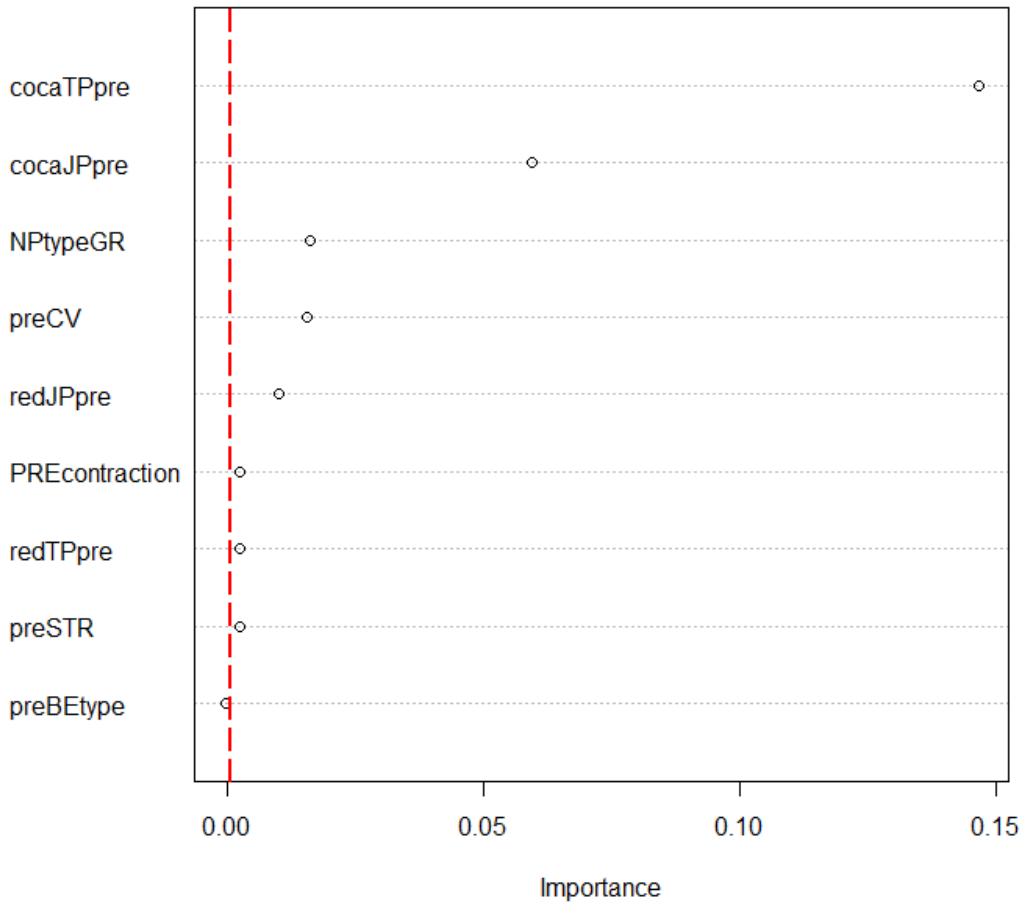


Figure 42. CDS preceding context factor variable importance ranking for contraction

For following context factors, as seen in Figure 43, probabilities based on the Redford Corpus frequencies perform better than probabilities based on COCA frequencies. Again, transitional probability performs better than joint probability, which is what we should expect: there is not such a strong bond between a contracted element and the word that follows it, but expected, accessible following words as measured through transitional probability, should allow a speaker to contract. It appears here that caregivers are sensitive to task frequencies and are contracting when a word predictable from task-centric context is the following word. The combination *is~'s totally* has a high probability based on Redford Corpus frequencies (-0.097) but lower based on COCA frequencies (-0.893). The combination *are~'re big* has a low probability based on Redford Corpus frequencies (-2.651), but a bit higher based on COCA frequencies (-2.128).

3.6.2.3. Multimodel Inferencing

Table 13 presents the output from the multimodel comparison of contraction in caregiver speech. Positive coefficients are associated with more contraction. High probability contexts are associated with more contraction, but preceding context has a stronger effect (cumulative probability = 1) than following context (cumulative probability = 0.85). This makes sense, as a contracted element cliticizes to the preceding element, not the following, forming a strong bond between the preceding element and the clitic. Speaker contraction rate is also predictive of contraction, which also makes sense. The variable of speaker contraction rate is calculated by looking at contractions and potential contractions of a variety of words, including the contraction of *BE*. Quartile is also significant for contraction. The second and third quartiles have higher contraction

rates than the first quartile, and the final quartile has significantly less. Construction also has a probable effect (cumulative probability = 0.91), with future and progressive constructions contracting more often than copula construction. Child Age had no real probable effect on contraction.

Following Context Variable Importance for CDS Contraction

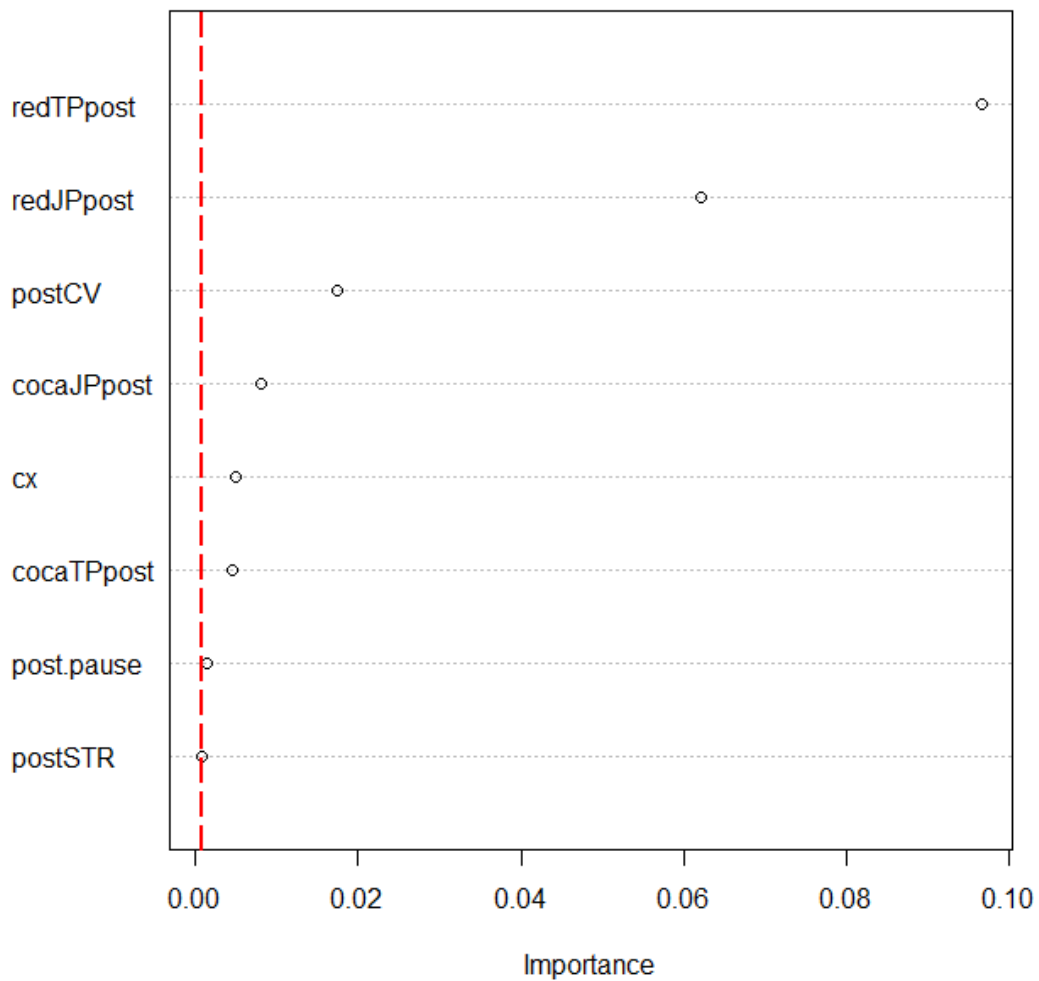


Figure 43. CDS following context factor variable importance ranking for contraction

Table 13. Multimodel Inferencing Output of CDS Contraction

Predictor	β^-	σ^-	LoCI	HiCI	p	CP
(Intercept)	1.731	0.118	1.500	1.963	0.000	NA
Preceding COCA TP	0.392	0.020	0.352	0.432	0.000	1
Speaker Contraction Rate	0.392	0.020	0.352	0.432	0.000	1
Quartile:						
Quartile 1 (Reference Level)						1
Quartile 2	0.263	0.036	0.192	0.333	0.000	
Quartile 3	0.083	0.041	0.002	0.163	0.044	
Quartile 4	-0.185	0.074	-0.330	-0.040	0.012	
Construction:						
Copula (Reference Level)						
Future	0.150	0.075	0.004	0.296	0.045	0.91
Passive	0.115	0.091	-0.062	0.293	0.202	
Progressive	0.061	0.035	-0.009	0.130	0.086	
Following COCA TP	-0.052	0.032	-0.115	0.012	0.109	0.85
Child Age	-0.001	0.001	-0.002	0.001	0.481	0.47

Note: reported values are coefficients with shrinkage and adjusted standard error, $n = 778$.

Table 14 presents the model comparison for the caregiver contraction model. We see fairly low model selection uncertainty, as the w for the best performing models equal 0.81, and only two models have Δ below 2. The second best model has all of the factors and the best model has all of the factors except for child age. In this case, Burnham and Anderson (2002: 131) say that the second best model with just one extra factor is not a serious contender. Therefore, there is very high certainty that the highest ranked model is the best one to account for the data with the given factors.

Table 14. Models of CDS Contraction with a Δ below 2.

Model factors	k	df	log likelihood	AICc	Δ	w
23456	5	13	-238.82	504.12	0	0.43
123456	6	14	-237.9	504.34	0.22	0.38
(Null)	0	4	-472.15	952.36	448.24	0

Note: Cutoff: $\Delta < 0.02$, 1 = Child Age, 2 = Following COCA TP, 3 = Preceding COCA TP, 4 = Speaker Contraction Rate, 5 = Construction, 6 = Utterance Quartile

3.6.2.4. Conclusion for BE Contraction in Child-Directed Speech

Now we return to the research questions put forth at the beginning of this section.

Q10. Are there differences in caregiver contraction depending on the age of the child?

Not really. There is only a very minor effect of child age on caregiver contraction.

Caregivers contract slightly less with older children, but the effect does not have enough cumulative probability in the multimodel inference output for us to take it seriously.

What matters more for adult contraction behavior is whether or not the specific adult is someone who tends to contract words (generally, not just the specific targets under investigation here) or not.

Q11. Are there different motivations for contraction vs. word shortening (duration reduction)?

Yes. Word shortening is influenced by following context probabilities, where high probability contexts result in shorter target words. Contraction, on the other hand, is influenced to a greater degree by preceding context probabilities, where high probability contexts result in contraction. Utterance quartile had no strong effect for adult word shortening (although there was an utterance final lengthening effect in child speech), but adults tend to contract more in the middle (second and third quartiles) of an utterance.

Additionally, there were no word shortening effects for *BE* by construction type, only for *HAVE*. In the case of *HAVE*, only one construction type is permissible to contract: the perfect, which also had a high degree of shortening. All *BE* construction types contract (for present tense inflections), and we saw that future and progressive constructions contract at a higher rate than copula or passive constructions in the adult data.

3.7. Contraction in Child Speech

As described in 3.13 more data was used for the analyses of contraction than for the analyses of phonetic dependent variables. The analyses that follow used 152 narrations from children as opposed to 74 that were used in the phonetic analyses. Because more data was used, it was possible to break up age groups into five different age groups (kindergartners v. first graders v. second graders v. third graders v. fourth and fifth graders) for random forest age comparison analyses rather than the four used in the previous sections. The main research questions for this section are:

Q12. Are there differences in contraction over time for children?

Q13. Are there different motivations for contraction vs. word shortening (duration reduction) in child speech?

3.7.1. CS Contraction of *HAVE*

As seen in Figure 44, most children, if they produce perfect auxiliaries, use full forms. Children who did not produce any perfect auxiliaries are not included in the figure. The children contracting perfect auxiliaries are all different children, not one child (nor one child over several time points) who happened to provide all contracted tokens.

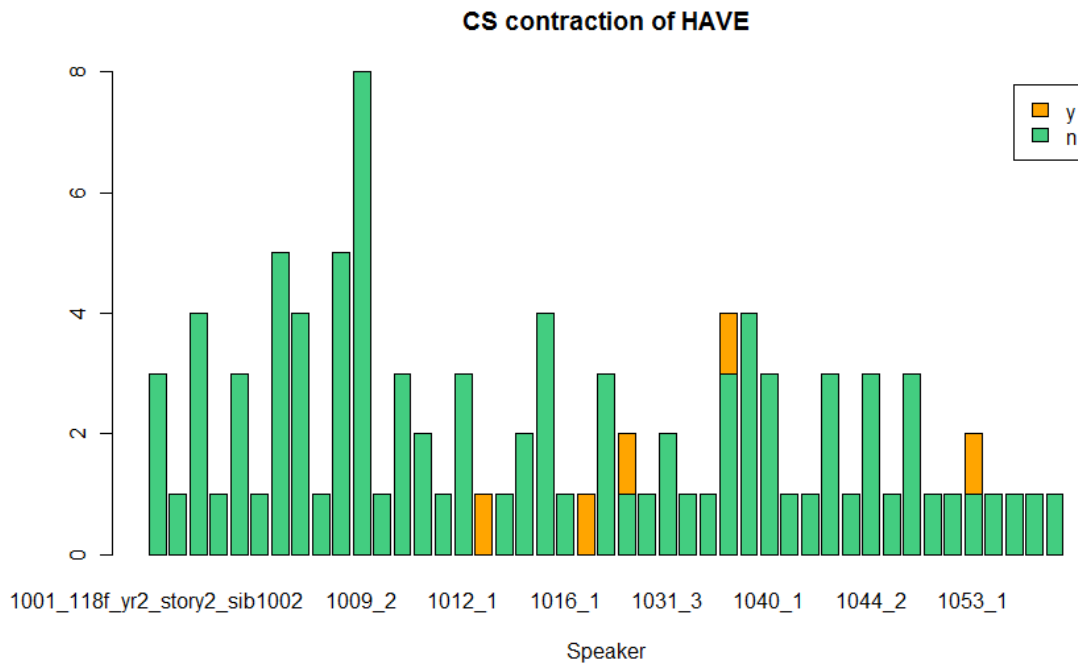


Figure 44. CS Contraction rate of Perfect *HAVE* by Speaker

As there are so few perfect auxiliary contractions, no analyses of *HAVE* contractions were done. However, there is a clear difference in child and caregiver behavior. Although caregivers contract perfect auxiliaries when speaking to their children, the children prefer the full form of the perfect auxiliary.

3.7.2. CS Contraction of *BE*

3.7.2.1. Introduction

As seen in Figure 45, some children never contract *BE*, but most contract far more than using full forms, especially the children who use many instances of *BE*.

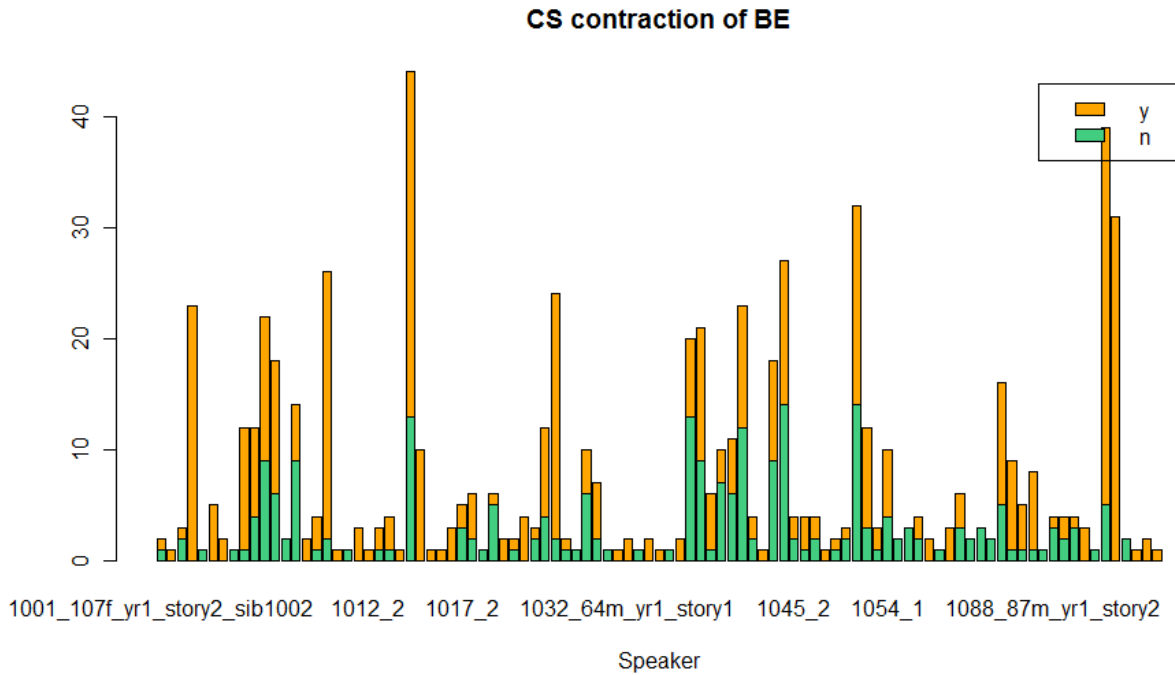


Figure 45. CS Contraction rate of Perfect *BE* by Speaker

3.7.2.2. Random Forest Variable Importance Ranking

This section presents random forest analysis of contraction, investigating speaker and word specific factors, preceding context factors and following context factors. Figures 46 and 47 show that, as with caregivers, speaker contraction rate is highly important (cf. Figure 41). Speaker (Gross) is ranked higher than Speaker (by text), indicating relatively high stability in individual differences between speakers over time. The utterance quartile that the target is in is also important, as is a speaker’s syllable rate and age. The particular inflection of *BE* is also ranked as important, indicating that some inflections of *BE* are more likely to contract than others.

Speaker Variable Importance for CS Contraction

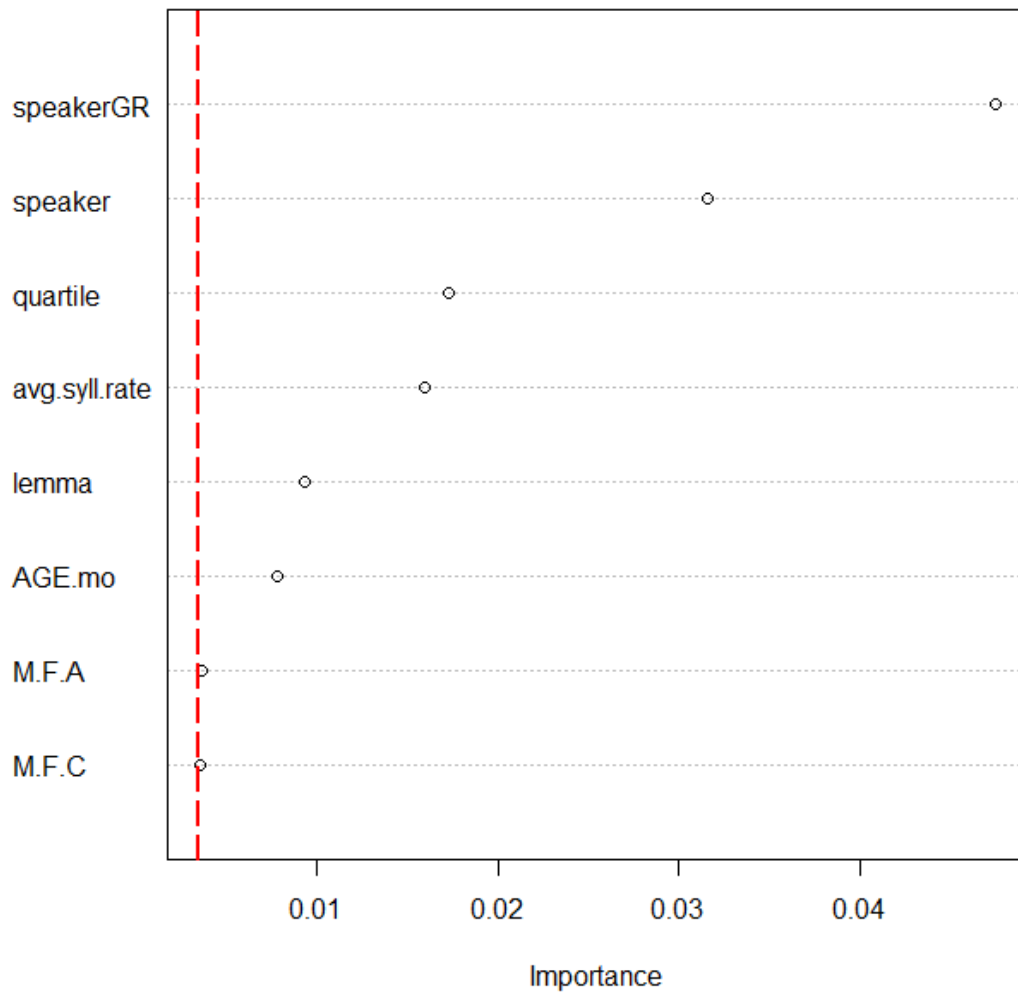


Figure 46. CS word and speaker specific factors variable importance ranking for contraction

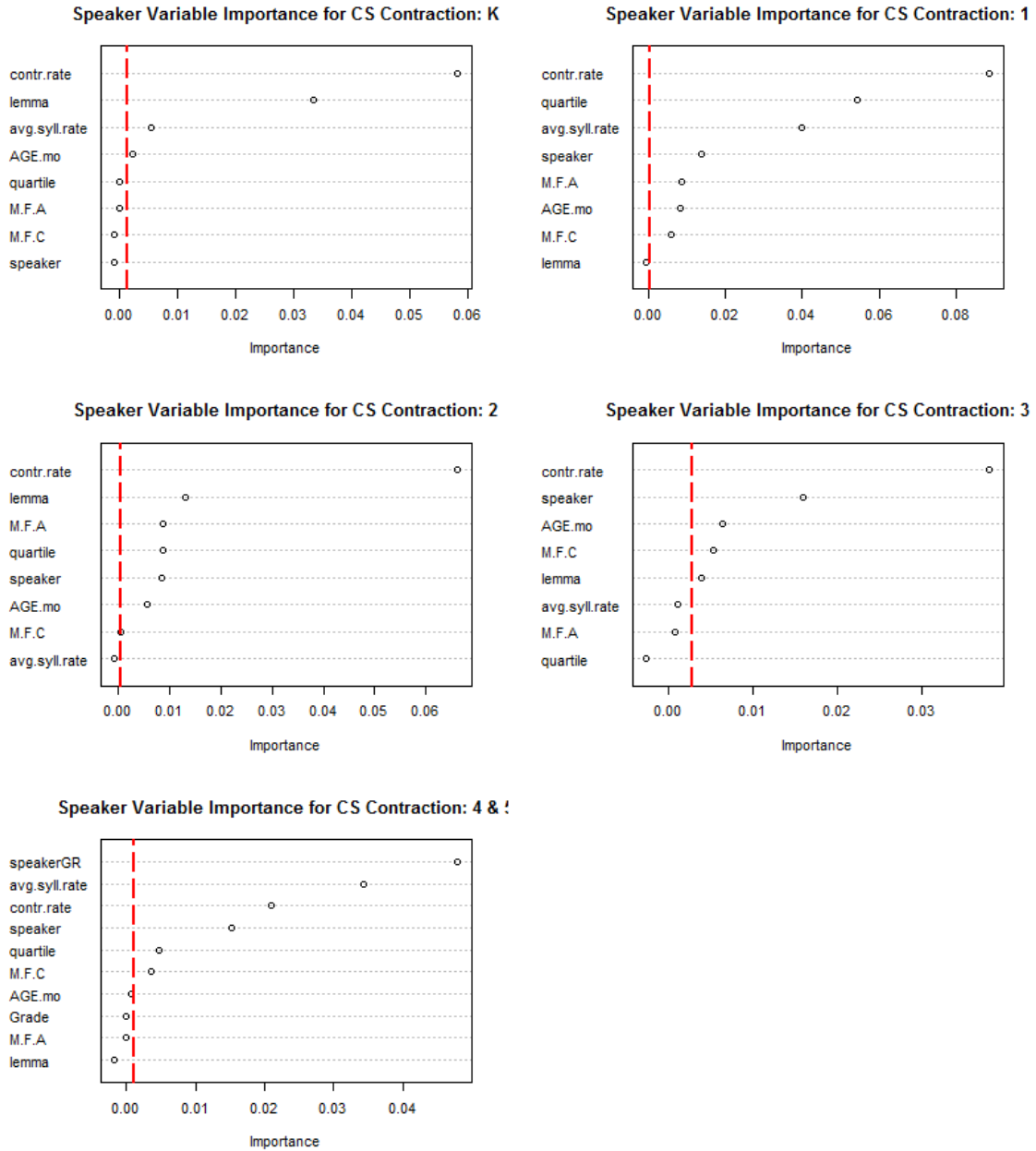


Figure 47. CS word and speaker specific factors variable importance ranking for contraction by grade

The preceding context factors importance ranking for children parallels the importance ranking for caregivers, as seen in Figure 48. The probabilities based on the COCA corpus perform best, and transitional probability performs better than joint probability. The priming variables of preceding *BE* type and preceding contraction are also ranked as important. As can be seen in Figure 49 below, these variables are not

influential in the younger groups, nor in the oldest group, which has a lower number of tokens. The data from the third graders is driving the high ranking of the priming effect.

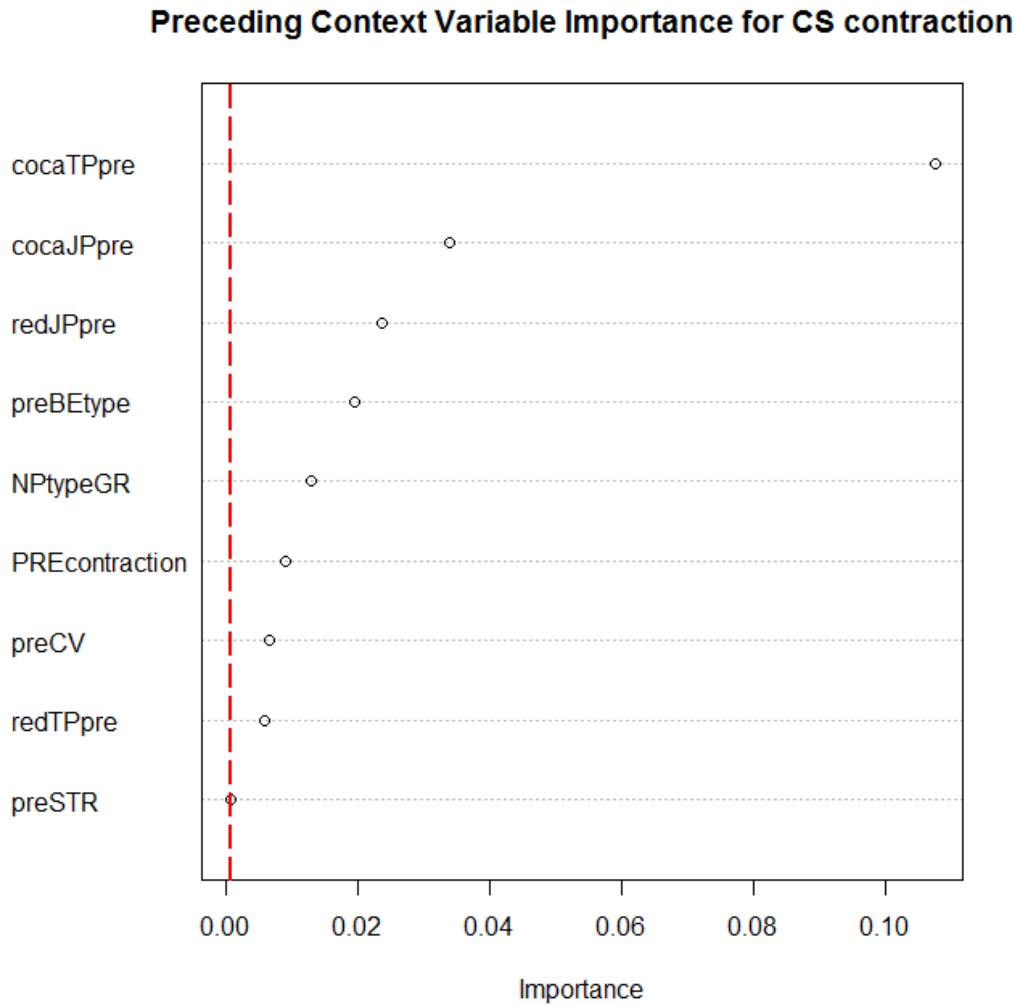


Figure 48. CS preceding context variable importance ranking of contraction

For the youngest group, probabilities based on the Redford Corpus frequencies are more influential than those based on the COCA frequencies, as seen in Figure 49.

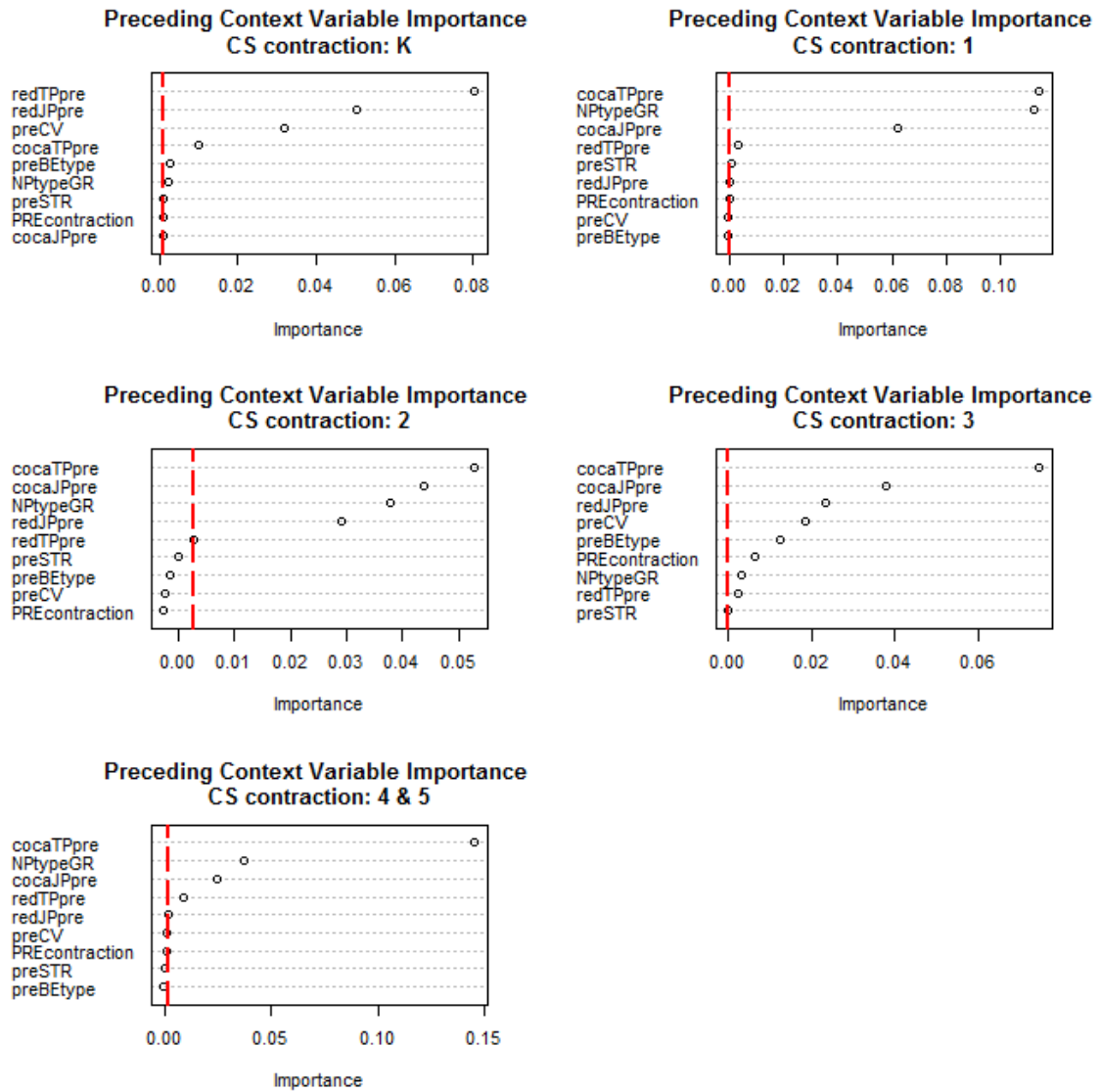


Figure 49. CS preceding context variable importance ranking of contraction by grade

Figure 50 shows that in the child data, as with the caregiver data, probabilities based on Redford Corpus frequencies are ranked as more important than probabilities based on COCA frequencies. This indicates that the children are also sensitive to the task-based frequencies and it is influencing their propensity to contract. Unlike with the adults, joint probability is ranked slightly higher than transitional probability, even

though these variables did not correlate highly ($r = 0.03$). Figure 51 shows that there is a lot of variation in variable importance rankings between the different age groups.

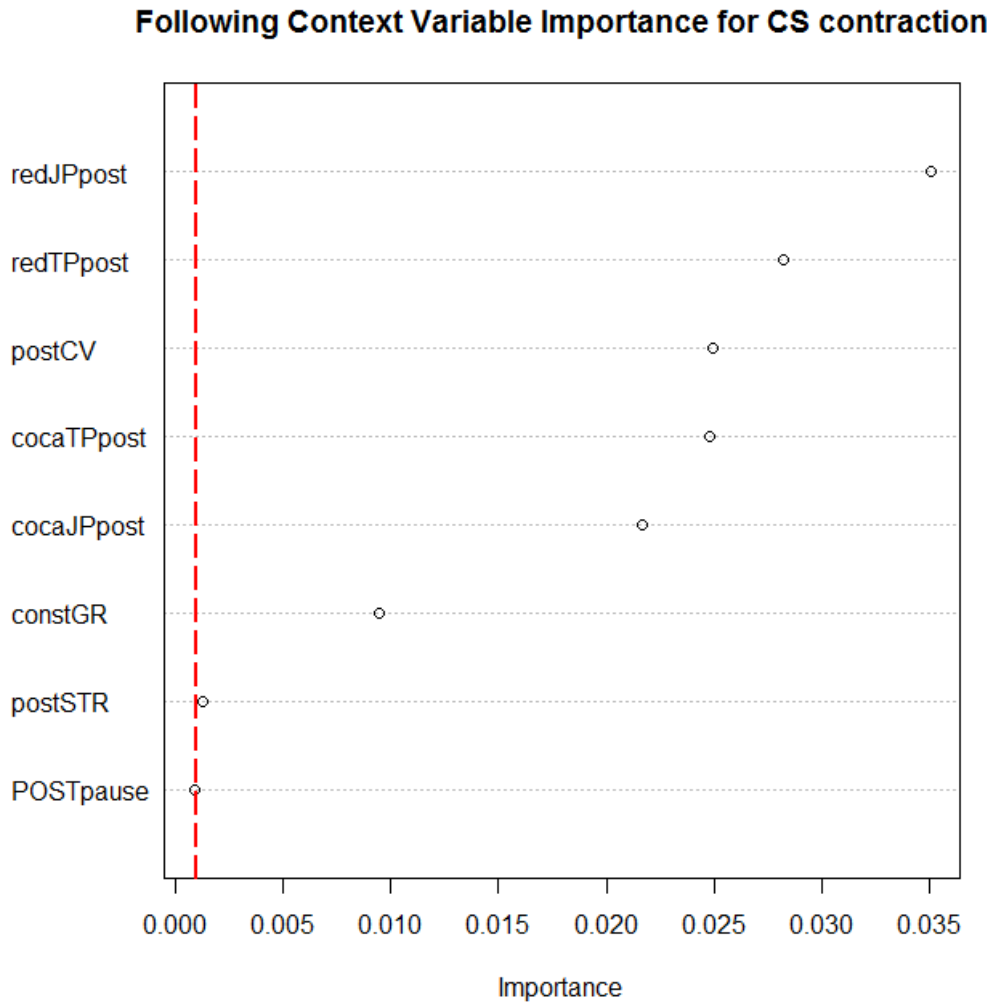


Figure 50. CS following context variable importance ranking of contraction

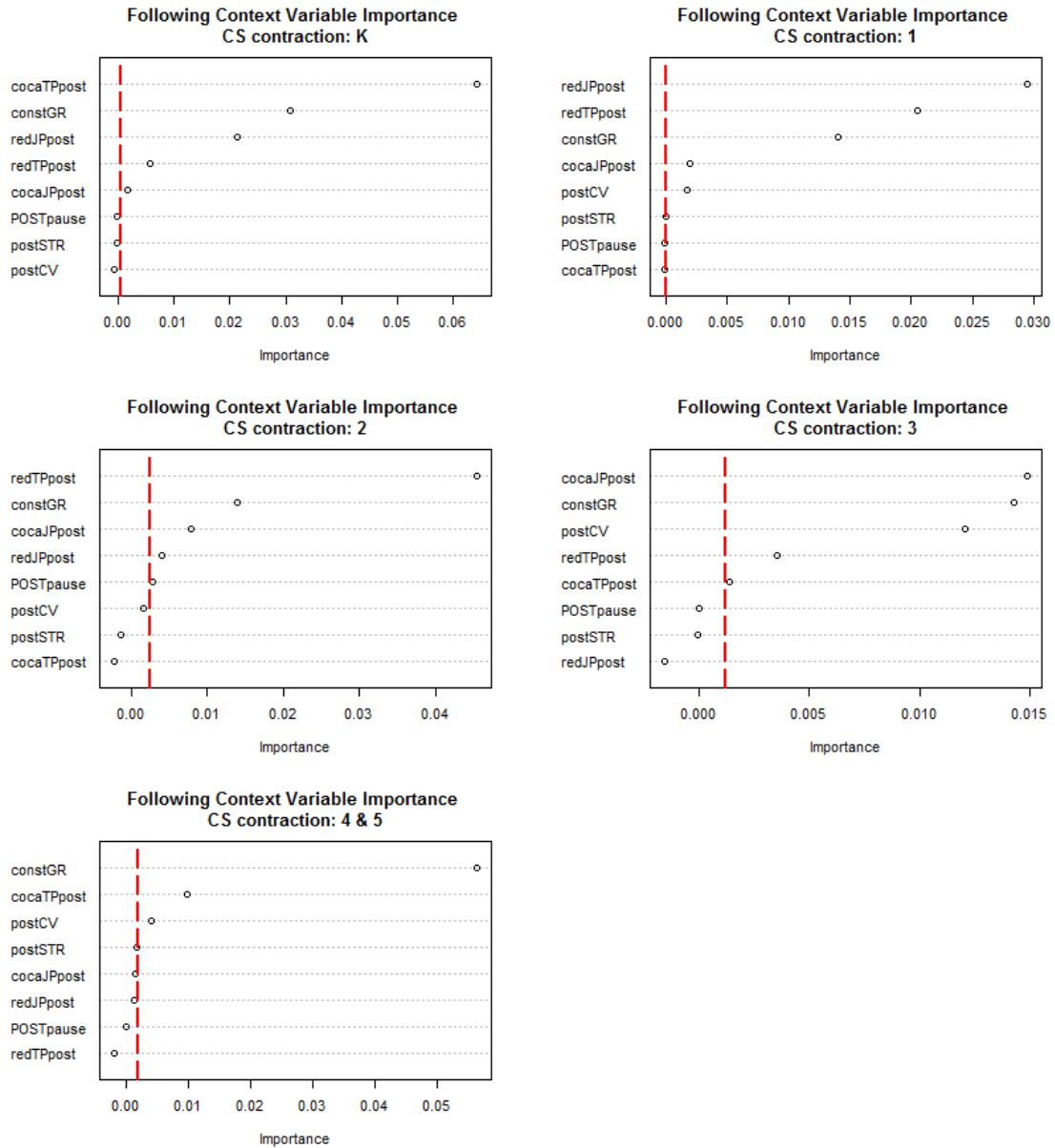


Figure 51. CS following context variable importance ranking of contraction by grade

3.7.2.3. Multimodel Inferencing

In Table 15 we see that several predictors have a very high cumulative probability, including interaction terms with Speaker Age. Positive coefficients are associated with higher contraction. As children get older, they contract more often

(cumulative probability = 1). When the target follows a highly probable context, it is also more likely to contract (cumulative probability = 1). In analyses of duration, the following context, rather than preceding context, was more predictive of shortened duration. As with caregivers, it makes sense for preceding context to have a stronger effect than following for contraction, as the contracted element cliticizes to the preceding element and there is no strong bond with the following element. There was no probable effect for following context probability and so it is not included in the output below. As with their caregivers, construction meaning is important for predicting contraction for children. Future auxiliaries contract more often than copulas, as do progressive auxiliaries. However, for children, passive auxiliaries contract less often than copulas. This result, however, is based on relatively few tokens of passive auxiliaries (copula: $n = 560$, progressive: $n = 165$, future: $n = 43$, passive: $n = 15$). There is also an interaction with age for construction. As children get older, they contract the passive auxiliary more and the future and progressive auxiliaries less. This goes in the opposite direction of the adult pattern. As with their caregivers, speaker contraction rate was highly predictive of contraction. People who have higher contraction rates contract *BE* forms more often than people with a lower contraction rate. There is an interaction with age, with a positive coefficient, meaning the facilitatory effect of high speaker contraction rate on target contraction is only strengthened over time. That is, an older child with a high contraction rate is even more likely to contract in a specific instance than a younger child with a high contraction rate. There is a high cumulative probability for utterance quartile, with the third quartile having less contraction than the first quartile, although there is no significant difference for the other quartiles. As seen in Figure 52, this does not capture

the real trends in the data. This figure shows that targets are much more likely to occur in the first half of the utterance (first two quartiles) and are much more likely to contract. Targets are successively less likely to occur in each of the next two quartiles and also have lower contraction rates in those quartiles.

Table 15. Multimodel Inferencing Output of CS Contraction

Predictor	β	σ	LoCI	HiCI	<i>p</i>	CP
(Intercept)	1.033	0.324	0.399	1.668	0.001	NA
Speaker Age	0.009	0.004	0.002	0.016	0.016	1
Preceding COCA JP	0.117	0.009	0.099	0.134	0.000	1
Construction:						
Copula (Reference Level)						
Future	0.787	0.477	-0.147	1.722	0.099	1
Passive	-4.733	2.255	-9.152	-0.314	0.036	
Progressive	0.832	0.170	0.498	1.166	0.000	
Speaker Contraction Rate	1.225	0.488	0.269	2.181	0.012	1
Speaker Age by Construction:						
Age by Copula (Reference Level)						
Age by Future	-0.007	0.005	-0.017	0.003	0.159	
Age by Passive	0.045	0.021	0.003	0.087	0.035	
Age by Progressive	-0.009	0.002	-0.013	-0.006	0.000	
Quartile:						
Quartile 1 (Reference Level)						
Quartile 2	0.018	0.035	-0.050	0.086	0.596	
Quartile 3	-0.106	0.046	-0.197	-0.015	0.022	
Quartile 4	-0.041	0.060	-0.158	0.076	0.492	
Speaker Age by Contraction Rate	-0.006	0.005	-0.017	0.005	0.274	0.62

Note: reported values are coefficients with shrinkage and adjusted standard error, *n* = 647.

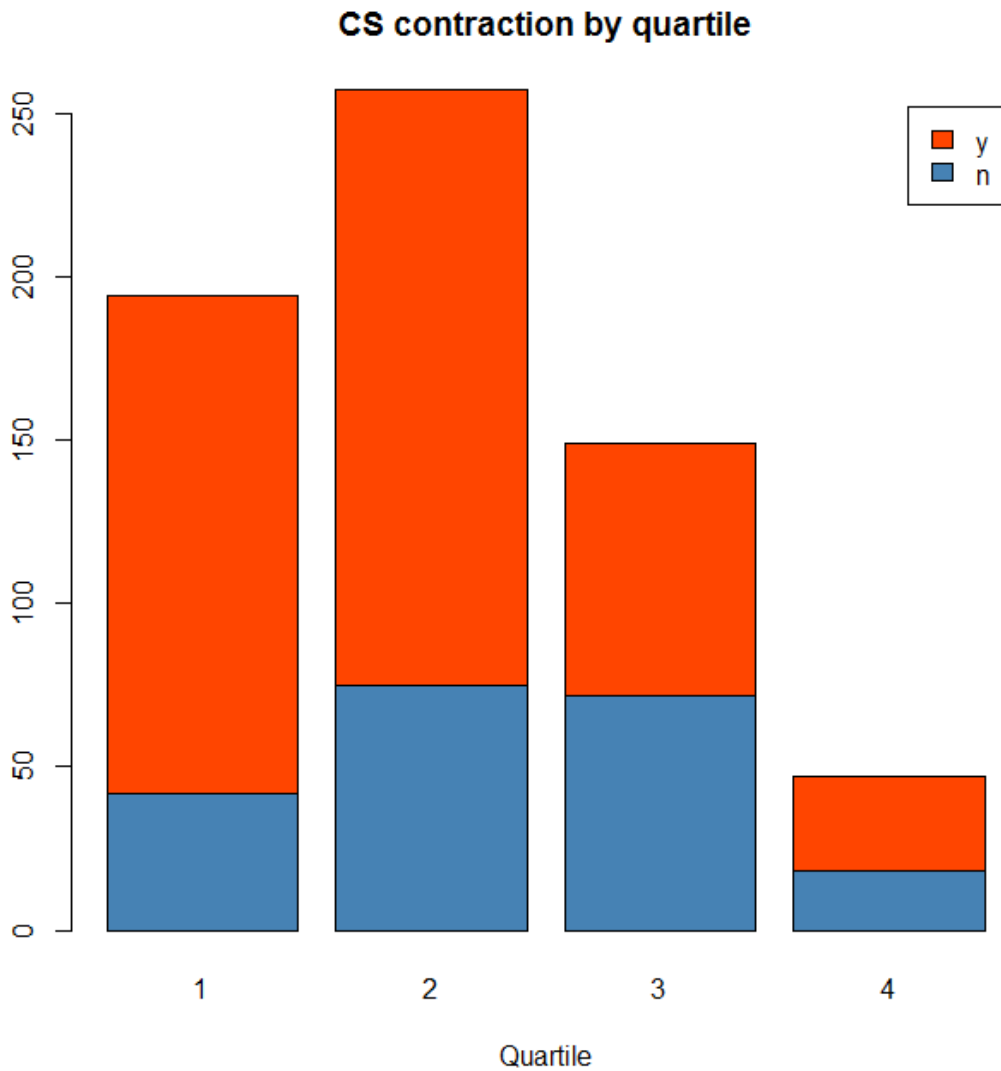


Figure 52. CS Contraction by Quartile

As seen in Table 16, the best performing model has a much higher weight than the next best performing model, and all of the best performing models have much lower AICc scores than the null model. This indicates rather low model selection uncertainty, with the top model being likely to be the most predictive model.

Table 16. Models of CS Contraction with a Δ below 2.

Model factors	k	df	log likelihood	AICc	Δ	w
1234567	7	17	-257.62	550.21	0	0.65
123456	6	16	-259.46	551.78	1.56	0.3
(Null)	0	4	-395.72	799.5	249.29	0

Note: Cutoff: $\Delta < 0.02$, 1 = Speaker Age, 2 = Preceding COCA JP, 3 = Construction, 4 = Speaker Contraction Rate, 5 = Utterance Quartile, 6 = Age by Construction, 7 = Speaker Age by Speaker Contraction Rate

3.7.2.4. Conclusion for BE Contraction in Child-Speech

Now we return to the research questions set forth at the beginning of this section.

Q12. Are there differences in contraction over time for children?

Yes. Children have high rates of contraction in the Redford Corpus, and they only contract more as they get older. There is also an interaction for construction type with age. Overall, as with caregivers, the progressive and future constructions are more likely to contract than the copula, and especially passive constructions. The effect for contraction in future constructions is stronger for adults and the effect for lack of contraction in passive constructions is stronger for children, but they show a similar pattern overall. However, as children get older, they contract progressive constructions a bit less and passive constructions a bit more. Speaker contraction rate has a facilitatory effect on contraction of a specific instance for all children and this effect strengthens as children get older. This indicates that the children follow their information compression strategy to a greater degree as they get older. If they go with the contraction strategy, they follow it more fully and to a greater degree than younger children.

Q13. Are there different motivations for contraction vs. word shortening (duration reduction) in child speech?

Somewhat. As with adults, construction type matters for *BE* contraction, even though it did not matter for *BE* word shortening. Although caregivers had more contraction mid-utterance and no word lengthening, children showed lengthening at the end of an utterance and more contraction early in an utterance. This pattern shows reduction (either of duration or by contraction) early and lack of reduction late in the utterance. Children, then, are more sensitive to position of the target word in the utterance than adults. Additionally, the targets under investigation here tend to occur earlier in the utterance rather than later. Children are likely to reduce targets when they are in their more probable utterance position (first half). This may indicate that children are also sensitive to utterance position probability in addition to bigram probability. When a target is occurring in an unlikely portion of the utterance, it is perhaps more difficult to process or access, leading to a lack of reduction, or they are letting their interlocutor know that something weird is going on: a verbal function word late in the utterance. For adults, it may be less of an issue due to a much better ability to plan long utterances which they may also expect of their interlocutors. For adults, who are more likely to produce complex noun phrase subjects than children, a 'late' auxiliary is less unusual.

3.8. Discussion and Conclusion

This chapter showed that over time children get better at signaling the unimportance of function words. As they get older, they reduce the proportion of time spent on function words, they contract more, they develop a strategy of information compression (shortening or contracting) and adhere more strongly to their strategy as they get older. In this section I will discuss types of reduction, information compression, child language proficiency and child-directed speech to school aged children. I will

conclude with a discussion of the statistical methodology used in this chapter, which will also be used in Chapter IV.

3.8.1. Results Summary

Based on the results above, children appear to be good at contracting and reducing the duration of words appropriately (in an adult-like way) even at the age of 5. Table 17 summarizes the results from this chapter.

For adults and children alike, probability is a very important factor in influencing reduction. High transitional probability contexts, particularly following contexts, are associated with shorter words and more contraction, for both children and adults. Perfect auxiliaries are shorter, for both children and adults. Children and adults alike seem aware of the more grammatical nature of the perfect auxiliary, as opposed to the lexical possessive and the semi-auxiliary modal verb which are always longer than the perfect auxiliary.

Following pauses and disfluencies are associated with lengthened word durations for both children and adults. Children also have more pauses and disfluencies than their caregivers. Based on my reflections of the data, following disfluencies are associated with problems of lexical access and high cognitive planning load (Bell et al., 2003; Bortfeld et al., 2001; Kapatsinski, 2010) for both children and adults, and for children, this is the case for preceding disfluencies as well. Pauses and disfluencies have been shown not only to be used for speakers to gain time to access a word, but also as a metalinguistic cue to their interlocutor that they are having trouble accessing a word and potentially seeking assistance in word recovery (Brennan and Schober, 2001; Schachter

et al., 1991; Schnadt, 2009). For adults, preceding pauses are associated with recovery and have no lengthening effect for word duration.

Table 17. Redford Corpus Results by Dependent and Independent Variables

	Normalized Duration	Contraction
Probability Variables		
High Following COCA TP	↓	↑
High Preceding COCA TP		↓
High Preceding COCA JP		↓
Construction		
Possessive	↑ ↑	
Modal	↑ ↑	
Perfect	↓ ↓	
Copula		↑ ↑
Passive		↑*
Progressive		↓ ↓*
Future		↓ ↓
Speaker Variables		
Age/Child Age	↓ ↓	↓
High Contraction Rate	↑ ↓*	↓ ↓*
Control Variables		
Preceding Stress	↓	
Following Pause	↑ ↑	
Quartile	↑	↑ ↑

Note: Green arrow indicates cumulative probability over 0.6 for caregiver speech, orange arrow indicates cumulative probability over 0.6 for child speech (factors with cumulative probability over 0.6 were in bold in the multimodel inferencing output tables). Arrows pointing up indicate the IV had a lengthening/non-reducing effect (or is associated with full forms for the contraction IV) and arrows pointing down indicate the IV had a shortening/reducing effect (or is associated with contracted forms for the contraction IV). An asterisk shows a interaction with age that reached a cumulative probability of 0.6.

Over the course of the study, adults decreased the length of grammatical words, as did children. Adults are signaling that grammatical words should be short and children are picking up on that and doing likewise.

3.8.2. Information Compression

One strong pattern we see in the data is that people have a reduction strategy for grammatical words: either they contract, or they reduce duration. Caregivers all do this, and children start to choose a strategy around age seven. Caregivers are fairly consistent in their contraction rates over the three years of the study (cf Figure 53, where only one speaker has variability in their contraction rate). Children have a lot of difference in their contraction rates over time (cf. Figure 54), but by age 7, they are a bit more stable (cf. Figure 55). However, they may still be testing out strategies even after age 7.

At age 7, children are not as consistent in their choice of information compression strategies as adults, but they have nonetheless made a leap forward from the younger children. Some children show more consistency in an information compression strategy than others, so individual differences likely also have an effect as well. Information compression is a kind of listener accommodation. Spending less time on unimportant words means speakers can spend more time on important words. A longer or clearer signal for more important words insures that a listener will be able to decode and understand those words more easily. Listeners do not need clear signals for expected or unimportant words because they are inferable from context. When children (or second language speakers) spend too much time on function words or frequent words they fail to accommodate their listeners by confusing them with strong signal strength for unimportant/predictable content (see Caballero and Kapatsinski, 2014, for evidence that

augmenting predictable information is unhelpful to the listener). Children who are more proficient speakers are the ones more likely to use an information compression strategy and be more consistent with its use early on. The next section examines issues of child language proficiency.

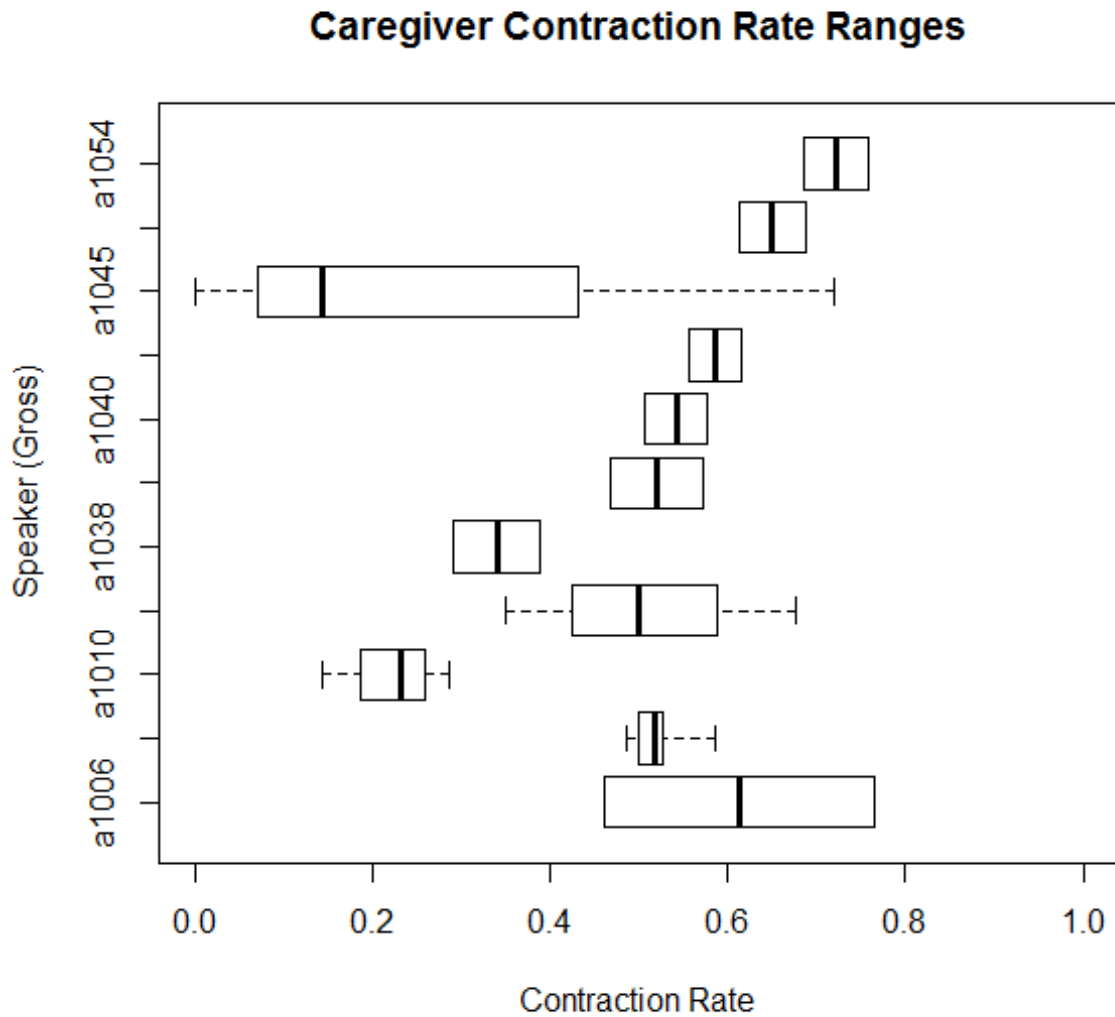


Figure 53. Caregiver contraction rates by speaker

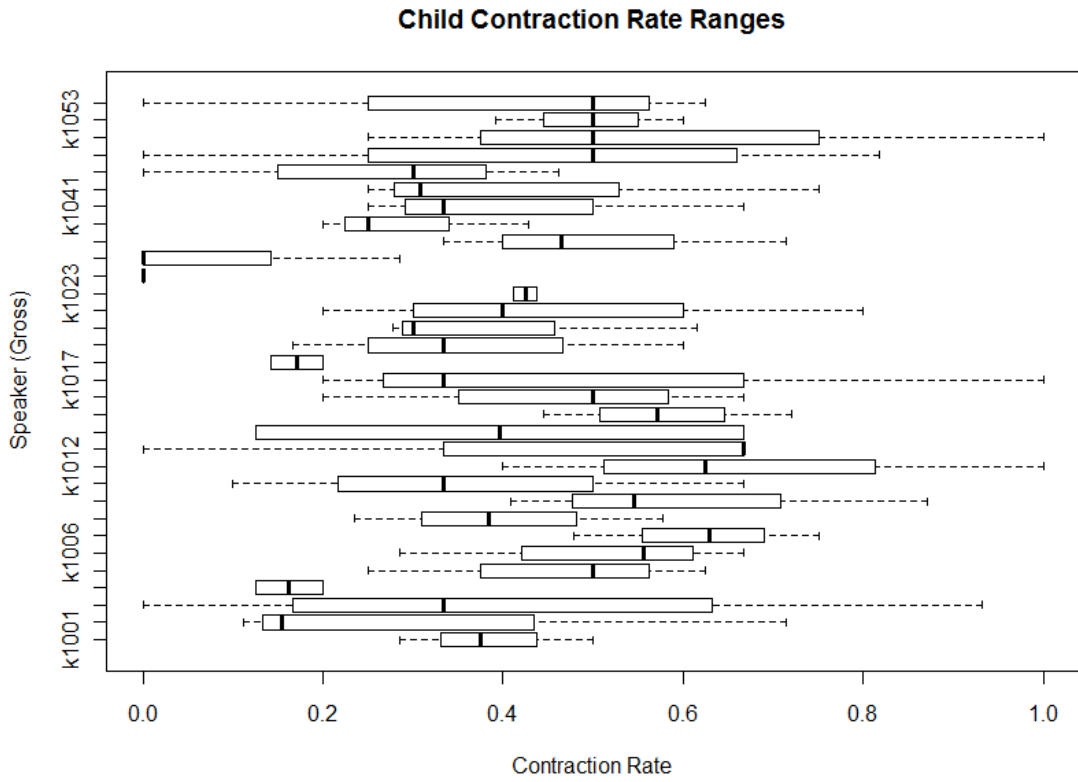


Figure 54. Child contraction rates over time by speaker

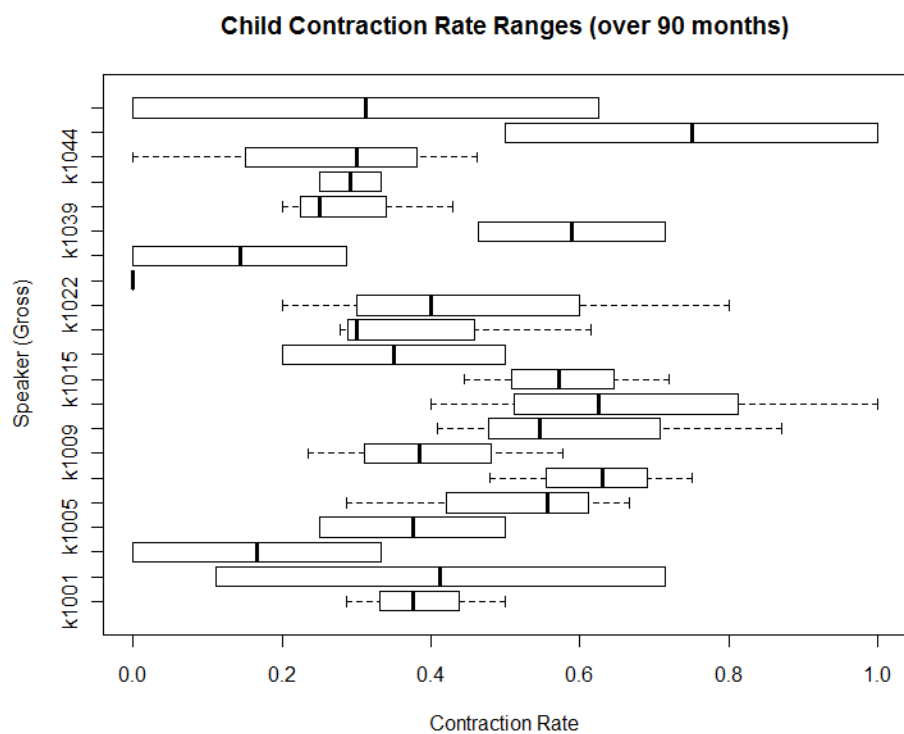


Figure 55. 7+ Child contraction rates over time by speaker

3.8.3. Child Language Proficiency

There are three measures, or indicators, of proficiency that are discussed in this chapter and they all have a relationship with child age and with reduction in child speech. The first measure is syllable rate. Children with a faster syllable rate have more articulatory control than children with a slower rate. They show an advanced *physical* ability. However, when they are able to speak faster, they in turn are also able to say more within in the same time period, and they do. Utterances do not decrease dramatically in milliseconds for older speakers; they are simply able to say more within a similar amount of time. This points toward advancement in *cognitive* abilities as well (Redford, 2014). The relationship between age and syllable rate is monotonic. As children get older, their syllable rates increase. When judging their effect on the

normalized auxiliary duration, which is a way of showing children's sensitivity to word importance and sensitivity to the needs of their listener, we see that generally age and syllable rate affect normalized duration in the same way. As shown in Figure 56, younger children have normalized duration proportions that are not strongly affected by syllable rates. For older children, however, faster speakers show longer normalized durations (red) than slower speakers (green). As we will see, this unexpected effect is due to an interaction with other means of evaluating proficiency in child speech.

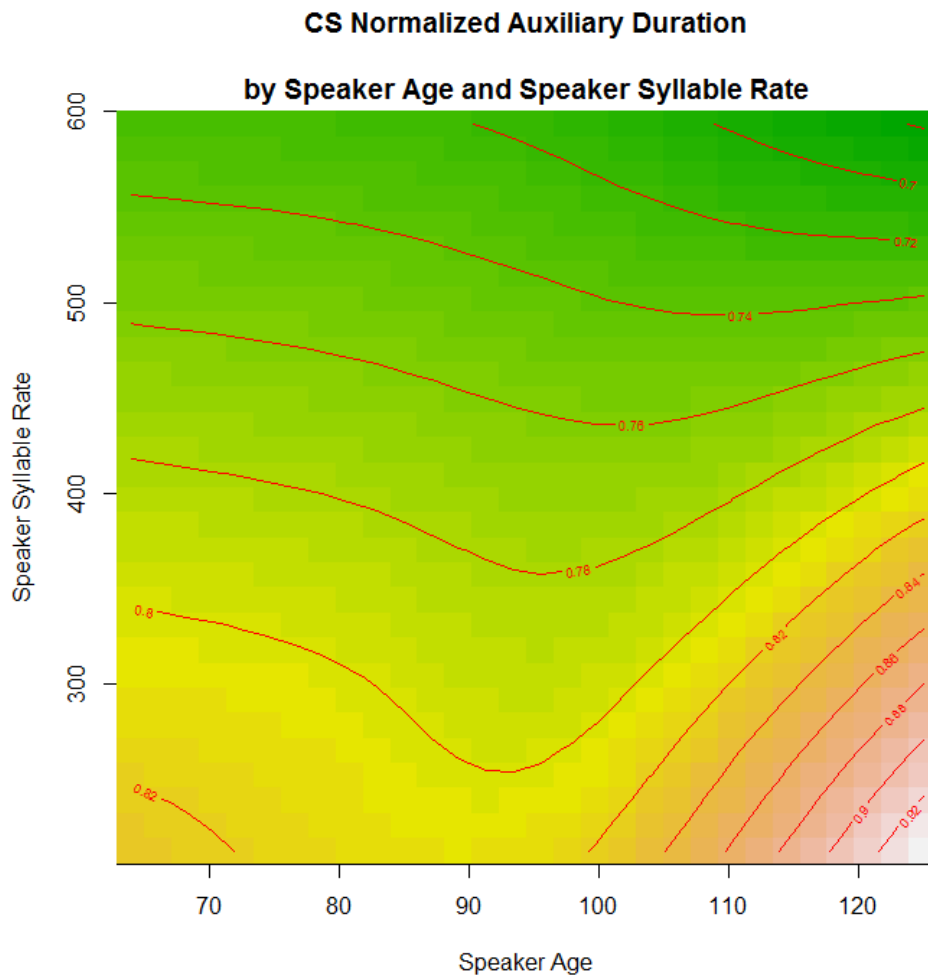


Figure 56. Relationship of Speaker Age and Syllable Rate on Duration Reduction in Child Speech

The second indicator of child language proficiency is text entropy. Entropy, as calculated here, is a vocabulary measure. Children use higher-level vocabulary words (rarer words) as they get older. All adults have higher text entropy than the children. Text entropy indicates a kind of *cognitive* ability, in word-learning and word choice skills. As shown in Figure 57, young children who have low text entropy scores also tend to have longer normalized auxiliary durations (red) and young children who have high text entropy scores tend to have shorter normalized auxiliary durations (green). This shows that young children who have good vocabularies also tend to be more sensitive to word importance and listener accommodation. For the oldest speakers, there is again a reversal, where older children who have very low entropy rates also have shorter normalized auxiliary durations (green) and older children who have very high entropy rates actually have longer normalized auxiliary durations (red).

The third measure of proficiency is information compression, which we saw above can be accomplished with short normalized durations or an increase in contractions (cf Figure 35 in section 3.5). On a listener-oriented account of reduction, information compression is a *social competence* measure, since it is predicted listener accommodation. When we examine the relationship between contraction rate and the other measures of proficiency (text entropy and syllable rate) absent of age, the picture becomes clearer. Figure 58 shows that children with high entropy rates tend to either have low contraction rates and reduce word durations a great deal (green in top left corner) or have high contraction rates and then do not reduce word durations (red in top right corner). It is children with low text entropy rates (less proficient on the vocabulary measure) that follow an all or nothing pattern, reducing and contracting to a high extent

(green in bottom right corner) or doing neither (red in bottom left corner). Figure 59 below shows essentially the same pattern for syllable rate as Figure 35 shows for age.

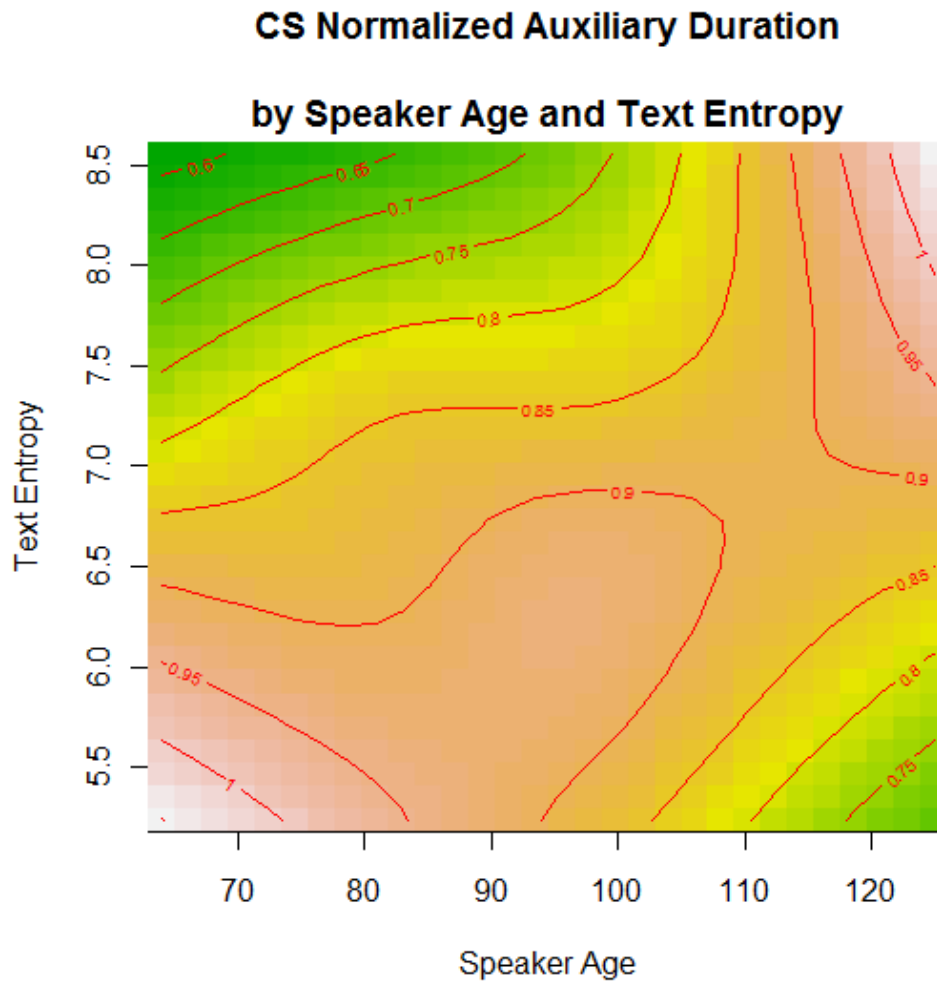


Figure 57. Relationship of Speaker Age and Text Entropy on Duration Reduction in Child Speech

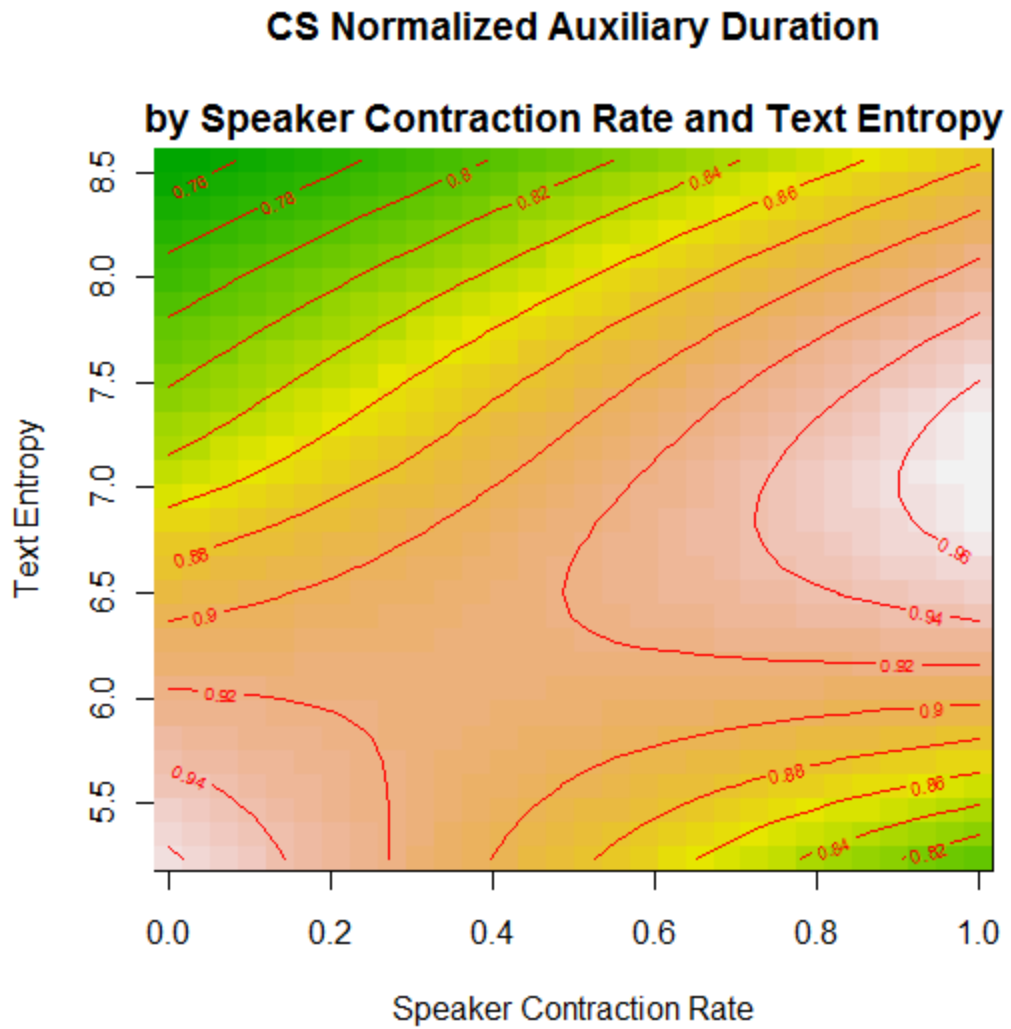


Figure 58. Relationship of Contraction Rate and Text Entropy on Duration Reduction in Child Speech

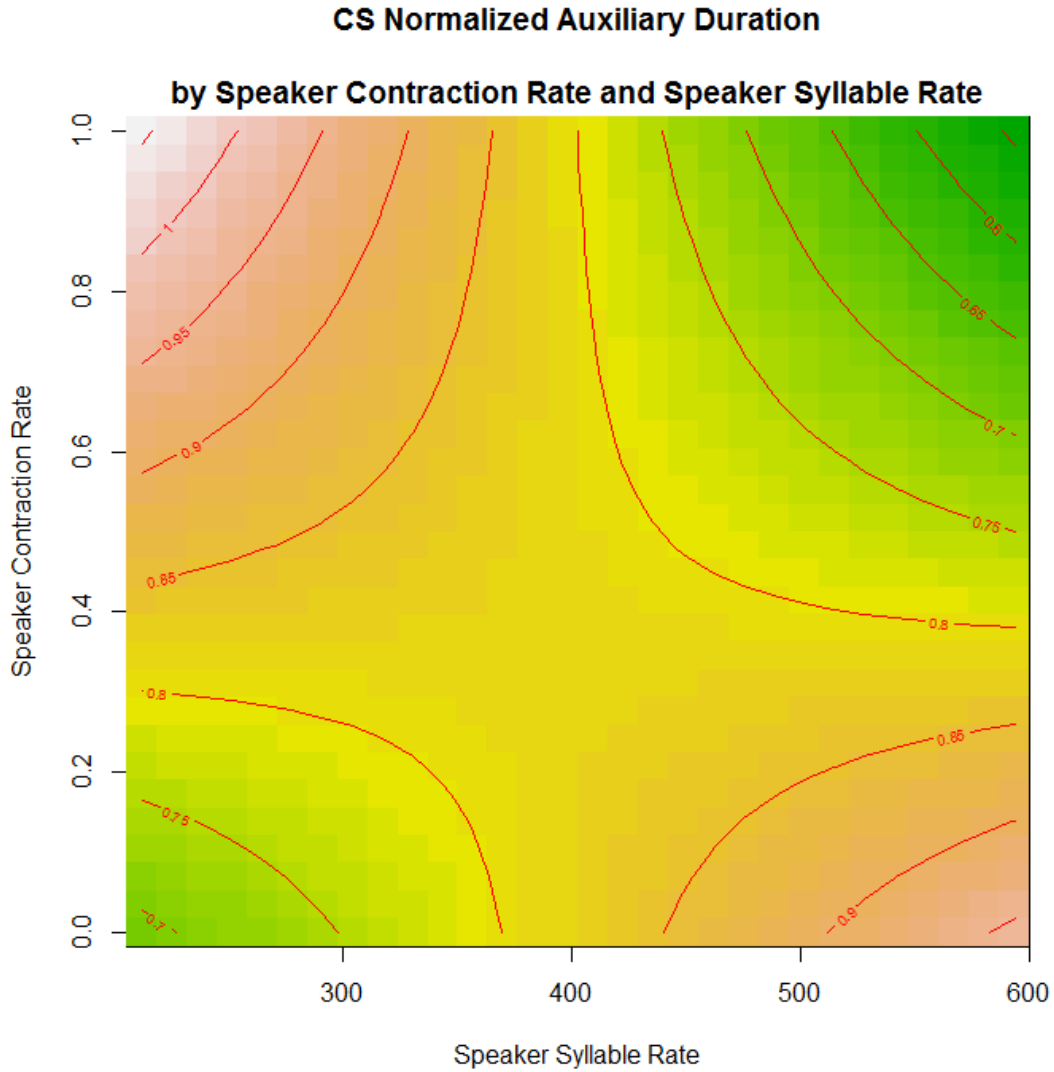


Figure 59. Relationship of Contraction Rate and Syllable Rate on Duration Reduction in Child Speech

When all of these variables are taken together (syllable rate for physical ability, text entropy for cognitive ability, the effect of informativity on reduction for social competence), we see that children who are proficient in one area tend to be proficient in other areas and this is not only due to their age. There are at least a few older children who show low proficiency on a couple measures, as indicated by the pattern reversals in Figures 57 and 56. When looking at the speaker details, it seems that three of the older children have low text entropy and low contraction rates, despite having fast syllable

rates. Another child has a medium contraction rate, but particularly low text entropy and also a very slow syllable rate. The effect of proficiency is stronger for young children on auxiliary duration. If a young child shows proficiency in one area, they tend to show it in other areas as well, particularly on the cognitive and social competence measures. There are several kindergartners and first graders in the study who have very high text entropy rates, and these children tend to contract a great deal, even while their syllable rates are still variable. Through these patterns, we see that certain children “get it” sooner than others and are able to show their language and social competence in a few different ways as a package. However, there are older children who are still developing in a few of these areas and need more time to acquire vocabulary and information compression skills. It may also be that these children choose not to perform well on this task in the lab. They may choose to take it easy (or perhaps rebel) and not show language competence through their word choice and not make an effort to accommodate their listener. In any case, these proficiency measures go together, either through ability or choice in performance.

3.8.4. Child Directed Speech Features for 5-10 Year-olds

Many of the features of CDS are no longer apparent in the speech of the caregivers in the Redford Corpus. As discussed in section 3.2.3, semantic content in the narratives is more complex for older children, but caregivers do not speak any slower with their younger children. Research of child directed speech (Swanson et al., 1992), which indicates that content words are lengthened and function words are not, meaning that there is a very low proportion of time spent on function words compared to content word. Based on this, I predicted that the duration proportion of function words would increase as children got older and caregivers no longer lengthened content words. The

opposite was true for the age group in the present study. Swanson et al. (1992) studied speech of caregivers (mothers) of children 1;6-2;4. The present results indicate that between the ages of 2 and 5, caregivers must stop lengthening content words and after the age of 5, caregivers start to reduce function words. Figure 60 below shows a schematization of this development. The utterances produced by caregivers get longer over this time period, but the proportions of content words to function words decreases and then increases again.

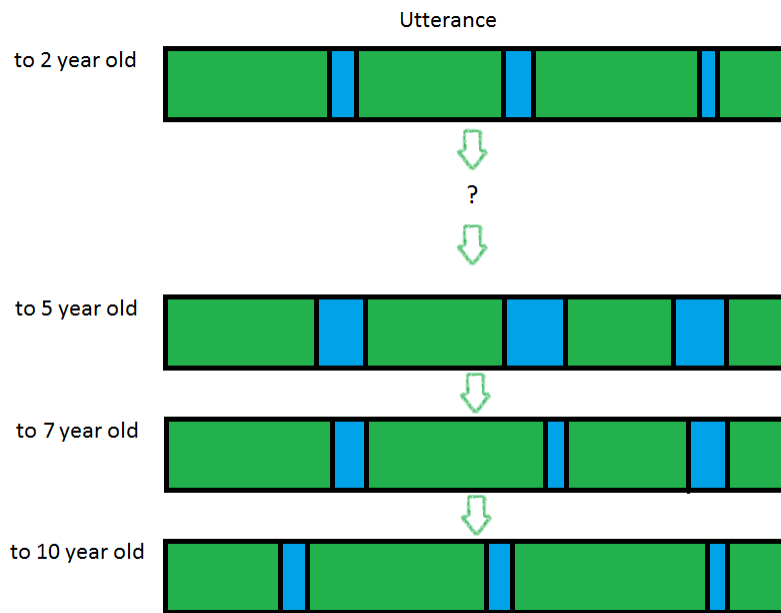


Figure 60. Schema of caregiver content to function word duration proportion

3.8.5. Methods Comparison

Both recursive partitioning through random forests and multimodel inferencing mixed effects regression models were used in this chapter to examine function word reduction behavior in a small corpus. The random forests were used primarily as a first pass to see what factors made an impact on the dependent variables. All of the factors

examined had a theoretical basis for being included, but clearly not all had a strong effect. Random forests were particularly useful in this case because so many of the potential variables were collinear and could not be included together in a regression analysis. Rather than a lot of model testing and model comparison to determine which of several collinear predictors was the best one to use, random forests were built which gave clear rankings of factor performance. However, the random forest variable importance ranking did not show the directionality of effects and counted many factors as important that did not have high cumulative probability in the multimodel inferencing output. For example, in Figure 23 we saw that contraction rate had an important effect for word duration, but my hypothesis was that a higher contraction rate would result in shorter word durations. In the regression analysis, we saw that this effect was actually in the opposite direction: a higher contraction rate is associated with longer word durations. Additionally, it is potentially confusing that the numbers on the x-axis of the variable importance rankings are only relative values and are not comparable to x-axis values in other importance ranking graphs. As the rankings are relative to the predictors included, the importance values are also relative and do not directly reflect effect size, so some predictors that are ranked very highly in one graph make actually be less important than predictors ranked as low in another graph with a different set of predictors. For example, in CDS word shortening analyses, Figures 25 and 24 show probability variables ranked far above the threshold for preceding context factors, and close to the threshold for following context factors, despite following context factors being significant and preceding context factors being non-significant in the regression models.

After factor selection and collinearity-testing, multimodel inferencing was then conducted to evaluate effect directionality, significance levels and predictiveness, while constraining predictors to have monotonic, linear effects. All possible models were ranked and the multimodel comparison showed that sometimes the best performing model (as indicated through AICc scores or log likelihood values) was not the only reasonable model. Sometimes the second best model included all factors but one and achieved similar data coverage. The multimodel comparison procedure, along with the cumulative probability ranking in the output showed that sometimes many factors were likely to have an effect on the DV, and several combinations of those factors would make a reasonable, high performing model. These are the cases in which multimodel inference is particularly valuable in allowing the researcher to avoid committing to a single model in the absence of strong evidence that that model is the most predictive one (see also Barth and Kapatsinski, In Press). At other times, only one combination of factors would be a reasonable choice for building an effective model. In these cases, a standard model selection procedure would be equally appropriate. Nonetheless, one advantage of multimodel inference is that it allows us to empirically determine the degree of model selection uncertainty and therefore to evaluate whether model selection can be safely performed as well as a way to still make valid inferences when a unique best model cannot be chosen (see also Burnham & Anderson, 2012).

Dilts (2013) uses both random forest analyses and LMER modelling and concludes that random forest analyses are helpful for determining dominant factors among correlated predictors and discovering interactions and non-linear relationships between factors. He argues that random forest models should be used in conjunction with

LMER modelling because random forest analyses lack the ability to account for the by-word and by-speaker variation than can be accounted for by random slopes in mixed effects modelling. He states that this is especially important for studies of reduction where people have different reduction strategies and words are subject to particular reduction patterns of their own. However, he makes it clear that factors with many levels in the random slopes can lead to over fitting (also cf. Barth and Kapatsinski, under review), and that it is not clear how many “levels can be included without limiting the generalizability of results” (p. 143). In a comparison of generalized linear modeling, generalized linear mixed effects modelling and conditional inferencing using random forests, Tagliamonte and Baayen (2012) conclude that the latter two are the more advantageous. Mixed effects modelling allows for the inclusion of random effects that result in better fit of the model to data, as well as allowing for increased confidence about making generalizations from the specific data to a larger population. Random forests allow for the comparison of similar phenomena that use different factor level configurations, all within one analysis. They argue that using classification trees allows the researcher to see how the most important predictors work together in a data set, sometimes producing complex interactions that could not be captured in regression modelling. Because it is computationally expensive and because currently conditional inferencing does not handle well factors with many levels (such as those one may use as a random effect), Tagliamonte and Baayen (2012), like Dilts (2013), advise the use of random forests for a preliminary view of the impact of the factors under investigation on the data.

In the present study, we saw that in using both random forest analyses for variable importance ranking and multimodel inferencing with linear mixed-effect regression models that the random forest analyses often consider a predictor important that does not turn out to be significant in the regression analysis. The regression analyses in the present study were always mixed effects model with a random effect of speaker and often a random effect of word inflection as well. The inclusion of these random effects allows us to account for variation that is due to word or speaker difference, variation that is less important to us theoretically than variation due to differences in levels of the fixed effects predictors. This effectively reduces the amount of variation that fixed effects can account for, making a larger effect size necessary for fixed effects to be significant. Additionally, using a multimodel averaging approach punishes (reduces) the factor coefficients. The multimodel inferencing approach means that a larger effect size is necessary for there to be statistical significance. However, like random forests, multimodel inference draws the focus of attention away from statistical significance given a single model and towards the predictiveness of a factor across models. However, ‘important’ predictors in random forests are often non-probable in the regression analysis. Why then use random forests at all? Aside from using random forests to select the best of several collinear factors (Schneider, 2014; Tagliamonte and Baayen, 2012), the random forests gave us a sense of the data before regression model building. For example, we saw that preceding TP and preceding JP often performed equally well, but that following TP often far outperformed following JP. We also saw that the difference between the two variations of speaker (speakers collapsed by year and speakers treated as separate individuals for each year they participated) was higher for children than for adults and that age differences were

more important for younger children than older children. If we were doing regression model comparisons, these patterns might not have been as obvious. Additionally, I presented random forests by grade for the child data. These figures allowed us to see the movement (or lack thereof) of factors over time for children. This gave us a more complete picture that an interaction term for age in the regression models, which gives us a better simplified summary picture. Both of these analysis types, when used in conjunction, provide us with a good sense of what is going on in the data. With language data, that is necessarily probabilistic and has variation that is driven by multiple similar processes, these techniques are helpful for the researcher and give a more complete picture for readers.

CHAPTER IV

REDUCTION IN INTER-ADULT SPEECH

4.1. Introduction

This study examines polysemous grammatical words in inter-adult speech, that is, adults being interviewed by other adults. As in the corpus studies of child and child directed speech, I examine productions of *am*, *are*, *had*, *has*, *have*, *is*, *was*, and *were* and contracted variants *'m*, *'re*, *'s*, *'d* and *'ve*. I find, as in the speech of children and caregivers, probable context, particularly following transitional probability, increases the likelihood of word duration reduction, as do certain construction meanings. This study uses a large corpus to confirm the results found in the smaller Redford Corpus.

4.2. Data and Methodology

4.2.1. Data

The Buckeye Corpus (Pitt et al., 2007) contains 40 interviews with native residents of Central Ohio. All speakers are white middle or working-class individuals. Speakers are balanced for age and gender in the corpus. Each interview lasts about an hour and speakers were told they were participating in a focus group about local issues. Sound files were force aligned and then hand-corrected by the corpus creators. Time stamps from the corpus data were used to calculate the durations used in the present study using *Python*. Tokens were limited to words occurring utterance medially and not occurring next to pauses or disfluencies such as *um*, *uh*, *er*, or other hesitations or word

re-starts. Therefore, pauses and disfluencies were not considered as independent variables for the Buckeye Corpus data.

4.2.2. Dependent Variables

4.2.2.1. *Normalized Duration*

Normalized duration was calculated in the same manner as in the other studies (the length of the word divided by the length of the average syllable in the utterance).

4.2.2.2. *Contraction*

Contraction was a discrete DV: contraction or not. Contraction was only examined in places where it was possible, so examples were eliminated that had an *is* following a word ending in a sibilant such as *forgiveness, phrase, lodge, Texas, this, which, etc.*

4.2.3. Independent Variables

4.2.3.1. *Probability and Construction Variables*

4.2.3.1.1. Probability

Tokens were coded for preceding and following context probability, both joint and transitional, using frequencies from the COCA (Davies 2008-). Too many of the words neighboring the targets did not occur in the Redford Corpus to be able to compare the effects of Redford Corpus frequencies. There are therefore four probability context variables to investigate: preceding joint, preceding transitional, following joint and following transitional probability. Probability measures are logarithmically transformed. Preceding transitional probability is a forward transitional probability, calculating how likely it is that *have, is, was, etc.* will occur after a particular word, given the frequency of

that particular word. Following transitional probability is a backward transitional probability, calculating how likely it is that *have, is, was*, etc. will occur before a particular word given the frequency of that particular word. Table 18 below shows bigrams with particularly high and low joint and transitional probabilities.

Table 18. Bigram probability examples from the Buckeye Corpus

Probability measure	High probability bigram	Numerical value	Low probability bigram	Numerical value
Preceding Joint	<i>it is</i>	-2.60673	<i>Krushchev is</i>	-8.65321
Following Joint	<i>is a</i>	-2.87876	<i>are crummy</i>	-8.65321
Preceding Transitional	<i>there is</i>	-0.23407	<i>whether was</i>	-4.82295
Following Transitional	<i>was born</i>	-0.06226	<i>am during</i>	-4.98528

4.2.3.1.2. Construction

BE can occur in four different constructions: copula, future, passive or progressive. *HAVE* can be modal, possessive or perfect. Construction is investigated to see if there are any construction-specific effects of reduction beyond the specific following joint or transitional probability of the target instances.

4.2.3.2. Speaker Variables

4.2.3.2.1. Age and Gender

The Buckeye Corpus is balanced for gender and age with two groups for each. There are twenty women and twenty men speakers in the corpus. There are twenty ‘younger’ speakers and twenty ‘older’ speakers. Actual ages are not provided. As I do not

expect any developmental differences between older and younger adults, age (as well as gender) is a control variable for inter-adult data.

4.2.3.2.2. Speaker

Speaker is included as a random effect in following the mixed effects regression models used in the multimodel inferencing procedure. Each speaker ($N=40$) is one level in the variable. Speaker is also included in the random forest analyses that follow to determine whether and how much variability exists between speakers.

4.2.3.2.3. Syllable and Contraction Rates

For speaker syllable rate, instead of averaging syllable rate from every utterance per speaker in the corpus (as was done with the data from the Redford Corpus) syllable rates were averaged only from utterances containing a target word. This provided 87 - 355 (mean 184.85) utterances to average per speaker, which is comparable or more than the total number of utterances for each text/speaker in the Redford Corpus (31 – 144 utterances, mean = 78 for caregivers).

All speakers in the Buckeye Corpus had very high contraction rates, as compared to speakers in the Redford Corpus. All 40 speakers from the Buckeye Corpus had contraction rates above 0.50. One particular contraction that contributed to the high rates of contraction was *you know* contracting to *yknow*. Including *yknow* as a contraction resulted in contraction rate ranges of 0.57 – 0.86, whereas without *yknow* rates were slightly lower (0.50 – 0.81). For speakers in the Redford Corpus, *yknow* was included as a contraction but there were very few speakers who used it and they did not use it often. This difference is one of genre and context (interview vs. story narration) and possibly also age. In any case, two different contraction rates were calculated in order to determine

which made more of a difference for reduction because contraction like *yknow* is in some ways substantively different than a contraction like *it's* because it is a tag question and is clearly treated as a unit. Only in its tag question meaning can it be contracted (**yknow it to be true*).

4.2.3.3. Control Variables

4.2.3.3.1. Quartile

Word duration may vary as a function of position in the utterance. Just as for the Redford Corpus data, targets were coded as occurring in the first, second, third or fourth quartile of the utterance as a factor (non-numeric) predictor. Quartile is a control variable in the present study. I am interested in the effects of probability and speaker-based variability that happens when utterance quartile is statistically controlled.

4.2.3.3.2. Stress Context

Stress context is a control variable in the present study. Targets were coded for whether they were preceded or followed by a stressed syllable, as speakers may be more likely to reduce after a stressed syllable (Echols and Newport, 1992; Davis et al., 2000; Jusczyk et al., 1999; Young, 1991).

4.2.3.3.3. Phonological Context

Phonological context is a control variable in the present study. Targets were coded for whether they were preceded for followed by a consonant or vowel.

4.2.3.3.4. Subject Noun Phrase

Subject Noun Phrase Type is included as a control variable in the analyses below. This factor has three levels: personal pronoun (*I, you, he, she, it, they, we*) v. non-personal pronoun (e. g. *what, who, there, here*) v. nominal noun phrase (e. g. *my friend, a*

house, the neighbor, etc.). Subject Noun Phrase Type is expected to be highly correlated with the preceding probability context variables.

4.2.3.3.5. Priming

For lemma priming (cf. section 3.3.2.1), the previous ten words were searched for an occurrence of the target, rather than any inflection of *BE* (for *BE* targets) or *HAVE* (for *HAVE* targets) as was done for the Redford Corpus data. This different implementation of a repetition priming variable was easier to code for automatic processing in *Python*, a necessity for the large sample size of the Buckeye Corpus dataset.

4.2.4. Statistical Procedure

Some independent variables such as stress context and probability context were potentially collinear, so random forest variable importance ranking was done for preceding contexts, following contexts and speaker- and word-based variables using the *party()* package (Horton et al. 2006a, Horton et al. 2006b, Strobl et al. 2007, Strobl et al. 2008) in *R*. Variables that were ranked as important were tested for collinearity using the *Psych()* package (Revelle, 2014) in *R* and then non-collinear variables were tested for significance using multimodel inferencing using the *MuMin()* (Bartoń, 2013) and *lmerTest()* (Kuznetsova, 2014) packages in *R*. See section 3 for more details about this methodology.

4.3. Results

4.3.1. Duration

4.3.1.1. Word Distribution in Sample

Words with a normalized duration of below 0.19 or above 3.4 were removed from the sample leaving 7,394 tokens. The number of remaining tokens for each word is depicted in Table 19 below, as well as the ranges for normalized duration and raw duration in milliseconds.

Table 19. Word distribution and length ranges

Word	<i>n</i>	Range in milliseconds	Normalized range	Mean of normalized values
<i>am</i>	32	65.751 – 349.870	0.443 – 1.841	0.951
<i>are</i>	729	28.950 – 548.383	0.201 – 2.124	0.616
<i>had</i>	747	32.000 – 469.582	0.207 – 2.510	0.976
<i>has</i>	273	55.837 – 893.854	0.397 – 3.136	1.195
<i>have</i>	1,674	24.623 – 696.048	0.190 – 2.768	0.996
<i>is</i>	1,129	32.292 – 558.000	0.211 – 2.979	0.867
<i>was</i>	2,269	31.410 – 580.000	0.225 – 3.177	0.957
<i>were</i>	541	34.151 – 407.329	0.196 – 2.248	0.694

4.3.1.2. Random Forests

Random forests were built using the {party} package (Horton et al. 2006a, Horton et al. 2006b, Strobl et al. 2007, Strobl et al. 2008) in *R*. Potential variables were ranked in

three groups: speaker and word specific variables, preceding context variables and following context variables. Only variables ranked above the absolute value of the lowest ranked variable (red line) have any real importance. Variables ranked as important were checked for collinearity and non-collinear variables were then used in the multimodel analyses that follow. In the random forest analyses, all words (*BE* and *HAVE*) were collapsed and included.

The first variable group is word and speaker-specific variables, as depicted in the variable importance ranking in Figure 61. Quartile is included in this group of variables, as it is neither a preceding nor following context predictor. Word is by far the most important variable, and will therefore be included as a random factor in the multimodel analyses to follow. Speaker is also somewhat important and will also be included as a random factor. The next two highest ranked variables are only barely ranked as important: Quartile and Speaker Contraction Rate. These variables ended up being significant in the Redford Corpus dataset. In Figure 61 below we see the two different kinds of speaker contraction rate: *contr.rate* and *contr.rate.yknow*. Because the contraction rate including *yknow* is ranked higher than the one without, this indicates that for word shortening, more contraction means greater effect, and the kind of contraction does not matter.

Figure 62 shows the variable importance ranking for preceding context variables of duration in the Buckeye Corpus. The probability context variables are ranked highest, with Joint Probability outranking Transitional Probability. We saw this same ranking in child-directed speech (but not child speech) in the Redford Corpus dataset. Noun Phrase type, Preceding Stress and the priming variable Preceding Contraction were also ranked

as important. However, the priming variable examining a preceding occurrence of a target in the ten words preceding the target was not ranked as important.

Figure 63 shows the variable importance ranking for following context variables for duration using the Buckeye Corpus. Following Transitional Probability was ranked highly, just as it was in child-directed and child speech in the Redford Corpus. Construction was ranked highly, but its influence is increased by the difference between *BE* words and *HAVE* words which are part of different constructions. Following Stress Context is ranked as unimportant.

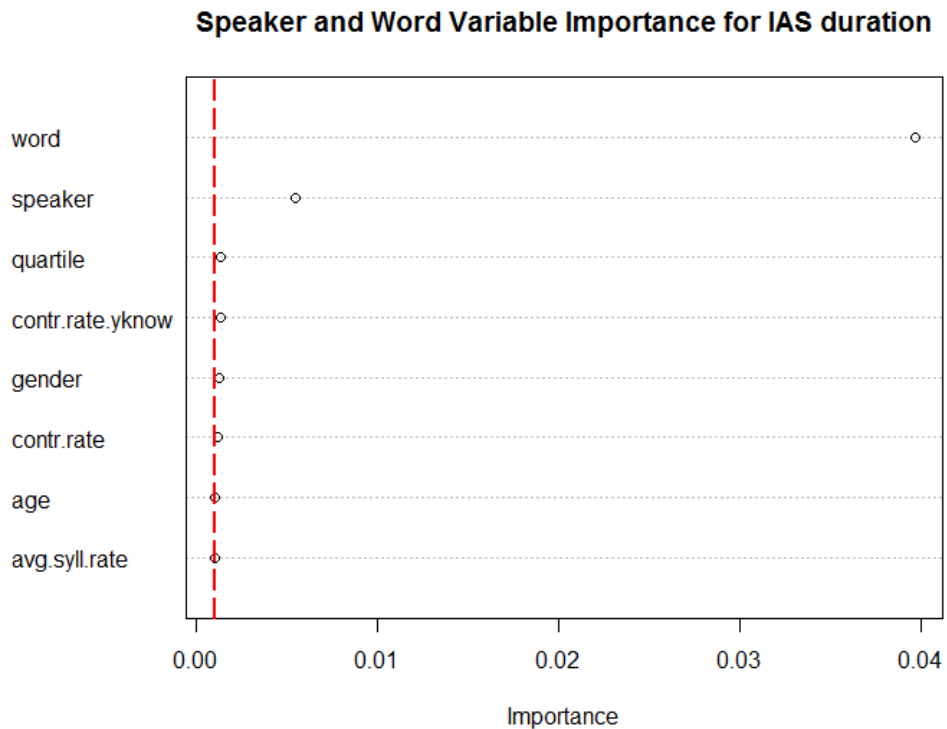


Figure 61. Random forest importance ranking of speaker-based and word-based factors for normalized word duration in IAS

Preceding Context Variable Importance for IAS duration

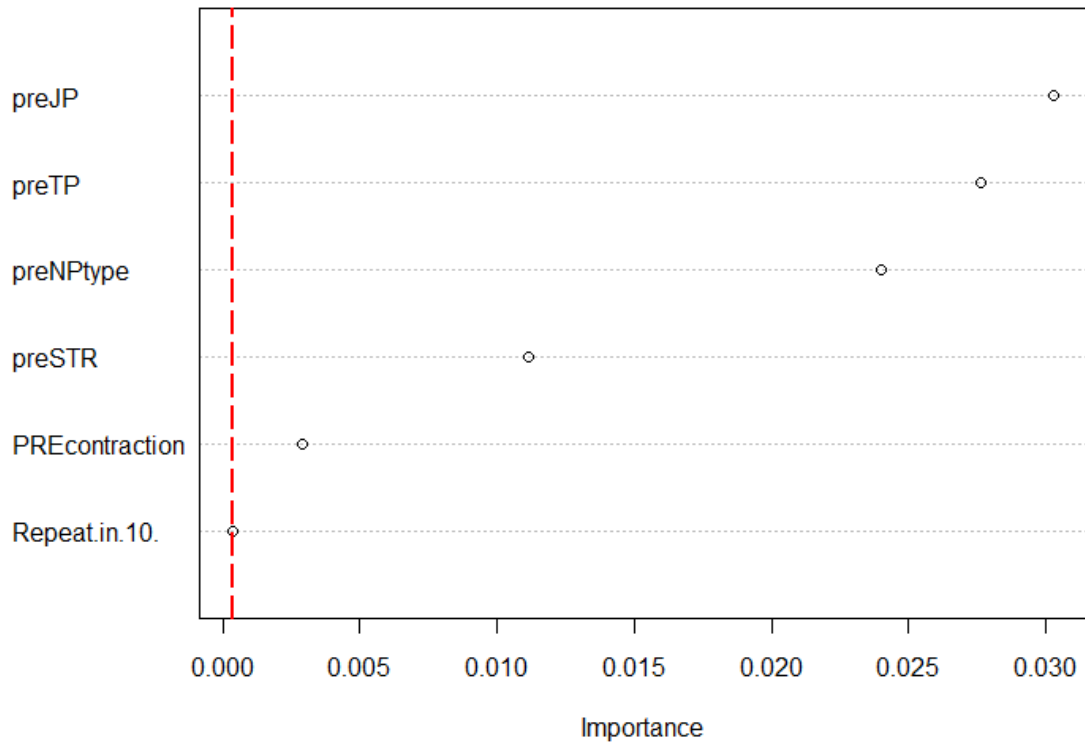


Figure 62. Random forest importance ranking of preceding context factors for normalized word duration in IAS

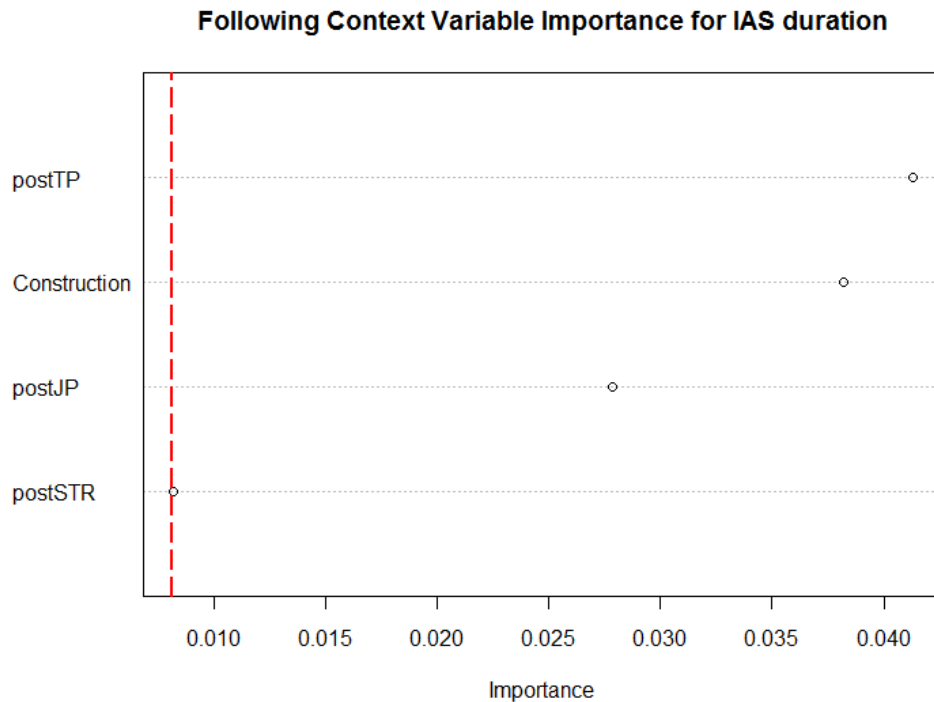


Figure 63. Random forest importance ranking of following context factors for normalized word duration in IAS

4.3.1.3. Multimodel Inferencing Full Model

Because we want to compare results of inter-adult speech to the results presented in Chapter III for child and child-directed speech, the main research questions for this section are:

Q14. What variables influence word shortening for adult speakers in a casual setting? Are these the same variables that influence word shortening for children and adults in a story narration task?

Q15. With the large Buckeye Corpus and a larger dataset, do variables reach significance that were not significant for the Redford Corpus data?

Multimodel inferencing was done for duration first on all words together then on inflections of *BE* and inflections of *HAVE*. Factors that were ranked as important, and were not collinear, were used to build a model and then a backward selection procedure was used until all variables had a cumulative probability of at least 0.40. Words in highly probable following contexts, as measured by transitional probability, had shorter normalized durations. Words that follow stressed syllables are shorter than words that follow unstressed syllables, as predicted, and as seen in child-directed speech. Higher joint probability, either preceding or following, results in longer word durations. This supports results seen in Bell et al. (2003), but not in the child-directed speech from the Redford Corpus. It may be that many data points are necessary before the lengthening effect of joint probability appears. Additionally, while not highly correlated in the Buckeye Corpus dataset, joint probability and transitional probability were often highly collinear in the Redford Corpus dataset, meaning the effects of joint probability could not always reliably be explored. The relationship between joint and transitional probability is explored further below. Next, there are two factors that show an effect of contraction on duration, and they tell a complex story. Speakers who contract more shorten their word durations more. This is the opposite effect of speaker contraction rate that we saw in child-directed speech and the speech of older children. However, the contraction rates for speakers in the Redford Corpus ranged from 0.0 to 1.0 for kids and 0.0 to 0.77 for caregivers. The contraction rates for speakers in the Buckeye Corpus are in a much narrower range, only 0.57 to 0.86. The range of contraction rates for these speakers may be too narrow to see the same contraction rate effect we saw in for the speakers in the Redford Corpus. However, when examining the 10 words before the target, if one of

those words is contracted, the target word is longer. This is consistent with the idea that speakers have a strategy to save effort on the production of grammatical words, but the effect is seen per utterance, rather than by speaker. Because speakers are contracting so often, the effect does not emerge from their specific contraction rates. However, for the speakers in the Redford Corpus, preceding contraction never was ranked highly in the random forest analyses for that data.

Table 20. IAS Duration Multimodel Inferencing Results

Predictor	β	σ	LoCI	HiCI	p	CP
(Intercept)	0.976	0.123	0.734	1.218	0.000	NA
Following Transitional Probability	-0.130	0.007	-0.144	-0.116	0.000	1
Preceding Contraction - Y	0.054	0.009	0.037	0.071	0.000	1
Preceding Stress Context - Y	-0.059	0.010	-0.078	-0.040	0.000	1
Preceding Joint Probability	0.042	0.007	0.028	0.057	0.000	1
Quartile:						
First (Reference Level)						1
Second	0.005	0.010	-0.014	0.023	0.632	
Third	0.018	0.010	-0.002	0.039	0.078	
Fourth	0.073	0.014	0.046	0.100	0.000	
Preceding Transitional Probability	0.056	0.029	-0.001	0.113	0.055	1
Pre JP * Pre TP	0.017	0.005	0.007	0.027	0.001	0.99
Following Joint Probability	0.007	0.005	-0.002	0.016	0.138	0.82
Speaker Contraction Rate	-0.135	0.139	-0.407	0.138	0.333	0.63

Note: reported values are coefficients with shrinkage and adjusted standard error, $n = 7394$.

Both models presented in Table 21 are reasonable models, containing following joint and transitional probability, preceding contraction, preceding stress context, utterance quartile, preceding joint probability and following joint probability and their interaction. Additional factors ranked as important in the random forest analyses all had cumulative probabilities under 0.50, showing they are not likely to make a substantial contribution to word duration difference. Speaker gender also made no substantial contribution to word duration difference in child-directed speech and were removed in the backwards elimination procedure.

Table 21. Models of IAS duration with a Δ below 2

Model factors	k	df	log likelihood	AICc	Δ	w
123456789	9	15	-2315.08	4660.22	0	0.52
23456789	8	14	-2316.63	4661.33	1.11	0.3
(Null)	0	4	-2591.05	5190.1	529.88	0

Note: Cutoff: $\Delta < 2$, 1 = Speaker Contraction Rate, 2 = Following JP, 3 = Following TP, 4 = Preceding Contraction, 5 = Preceding JP, 6 = Preceding Stress, 7 = Preceding TP, 8 = Quartile, 0 = Preceding JP * Preceding TP

Higher following transitional probability is associated with shorter durations but higher following joint probability is associated with longer durations. Why would this be the case? In Figure 64, we see that the lengthening effect (red area) of high joint probability seems to be restricted to cases where there is low transitional probability. Some specific bigrams that make up this area are *are and*, *was and* and *had that*. These cases have high joint probability as both words in the bigram are really frequent. However, they have low transitional probability because there is nothing about the word *and* or *that* that would lead a speaker to predict that any particular word would occur

before it. There are too many possibilities, so most bigrams containing *and* or *that* would have low transitional probability, even when they have high joint probability. We also see that, particularly for *are and* and *was and*, that these are generally cases of ellipsis, in sentences such as “*it doesn't really matter what sex they **are and** they treat them well*” or “*uh he was quite a bit younger than i **was and** i knew*” and even though there is high joint probability of the bigrams, these words do not form a grammatical unit. Targets that were cases of ellipsis were eliminated from the dataset for the Redford Corpus, so that may be another reason the lengthening effect of high joint probability was not found. In cases of *had that*, it would be possible that *that* is a relativizer, but for all the tokens measured, *that* was a determiner in utterances like “*yeah my roommate actually **had that** teacher and*”. For these cases, the lengthening cannot be due to ellipsis, but it is still the case that *that* is so general that it does not give clues as to what other words should occur with it in bigrams.

High Preceding Joint Probability and Transitional Probability also had overall lengthening effects. In Figure 65 below we see that the effect is a little complicated. There is generally not a strong effect of transitional probability or joint probability except for a sweet spot where both are high (although not in the very top range for joint probability), where there is shortening, indicated by a bright green area. The bigrams that make up this sweet spot are *I am*, *it is* and *it was*. The first two bigrams are incredibly likely to be contracted (as explored in section 4.3.2.3 below). These bigrams have both high JP and TP and are quite short.

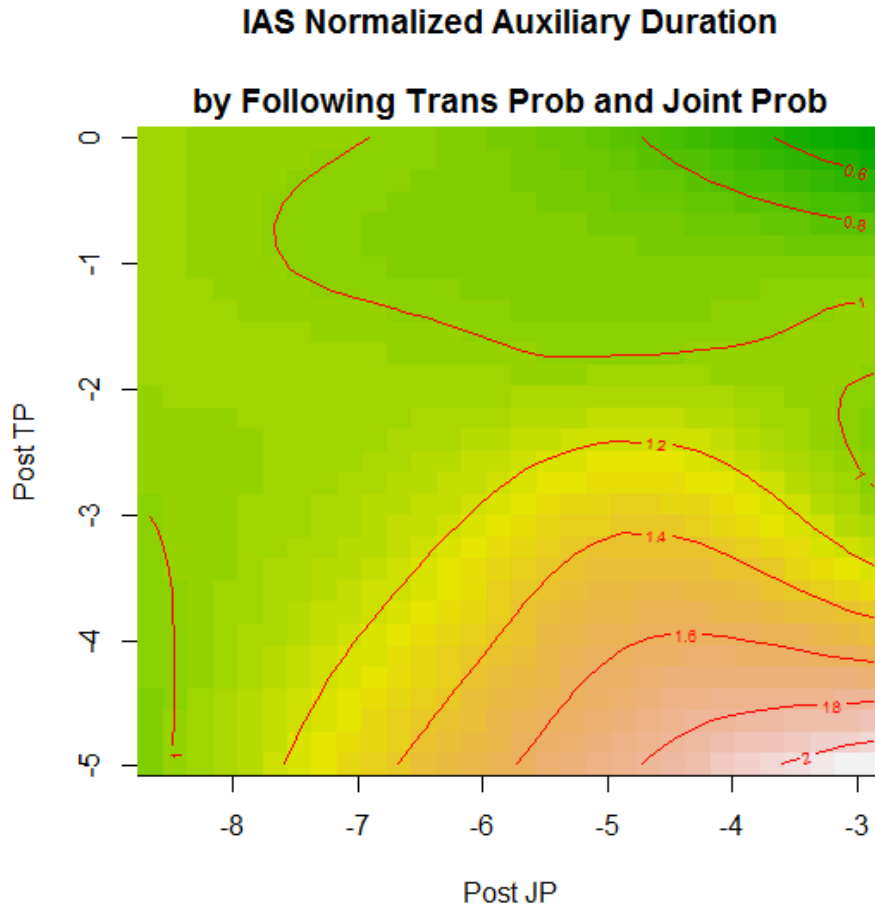


Figure 64. Following Joint Probability and Transitional Probability on Duration for IAS

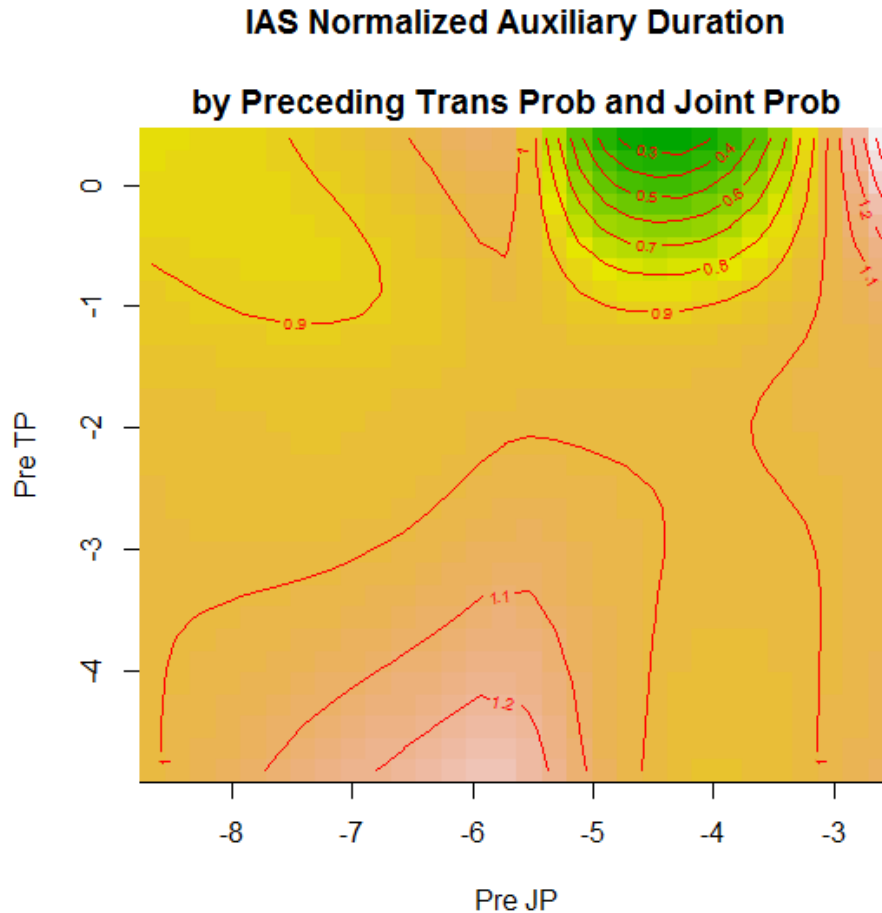


Figure 65. Preceding Joint Probability and Transitional Probability on Duration for IAS

When we return to our primary research questions for this section, we see that word shortening results from the Buckeye Corpus are remarkably similar to results from the Redford Corpus.

Q14. What variables influence word shortening for adult speakers in a casual setting? Are these the same variables that influence word shortening for children and adults in a story narration task?

Word shortening was primarily influenced by the same variables in both the Redford Corpus and Buckeye Corpus data, showing that results in the smaller dataset were

confirmed by the similar results from the larger dataset¹⁰. High Following Transitional Probability and the occurrence of a stressed syllable before the target both resulted in word shortening in child, child-directed and inter-adult speech. High Speaker Contraction Rate was a significant variable influencing the lengthening of function word duration in the Redford Corpus. In the Buckeye Corpus, high speaker contraction rate had a medium high cumulative probability (0.63) but did not reach significance. In the discussion above, this difference was attributed to the much higher rates of contraction for the speakers in the Buckeye Corpus, showing either a genre difference, audience difference (child-directed vs. adult-directed) or a speaker-specific differences. In the discussion of the Redford Corpus data, I argued that speakers were choosing an overall strategy for information compression: word shortening or contraction. If this is really the case, then everyone in the Buckeye Corpus has chosen contraction as their information compression strategy. However, we do see a significant lengthening effect for Preceding Contraction instead, an effect not found for the Redford Corpus data. If speakers have recently contracted, then they are likely to lengthen targets, showing a more local, rather than overall, information compression strategy. It is possible that with more data, this local information compression effect would have appeared in the Redford Corpus as well, particularly for speakers with high contraction rates.

Q15. With the large Buckeye Corpus and a larger dataset, do variables reach significance that were not significant for the Redford Corpus data?

¹⁰ Some variables, such as age and following pause, were only relevant for the Redford Corpus dataset. The speakers in the Buckeye Corpus were all adults, speaking to adults. Targets preceding or following a pause were eliminated from the Buckeye Corpus dataset.

Yes. Preceding context probability variables (Preceding TP and Preceding JP) were significant for the Buckeye Corpus dataset but not for the Redford Corpus dataset. Although preceding context variables were ranked as important in the random forest variable importance rankings, they did not end up being significant or probable in the multimodel regression analyses. It is possible that a larger dataset is needed for preceding context probability to have a strong enough effect. In the Buckeye corpus dataset we also saw significant lengthening in final quartile. Utterance quartile was ranked as probable, but not significant, for the child data in the Redford Corpus and as neither probable nor significant for the child-directed data in the Redford Corpus, despite being ranked as important in the random forest variable importance rankings. Again, it is possible that greater power is needed to see the effect of utterance final lengthening.

4.3.1.4. Multimodel Inferencing – Word Specific Results: BE

The primary research questions for this section are:

Q16. Are there construction specific effects for word shortening in the Buckeye Corpus data?

Q17. Are there word-specific effects for word shortening?

In order to answer these questions, I examine first the transcriptions of *BE* words as transcribed in the Buckeye Corpus and the distribution of those transcriptions. I then present a multimodel inferencing output for all *BE* instances in the corpus. No word-specific models were notably different from the collapsed *BE* model, and so no separate analyses are presented here.

In the Buckeye Corpus there are 86 different transcriptions of *was*. Figure 66 shows the distribution of transcriptions that occur 20 times or more. The citation form of [w ah z] is by far the most common, but reduced versions abound. Common reduced transcriptions show that often the initial glide is omitted, the final fricative is devoiced and other lax vowels are used.

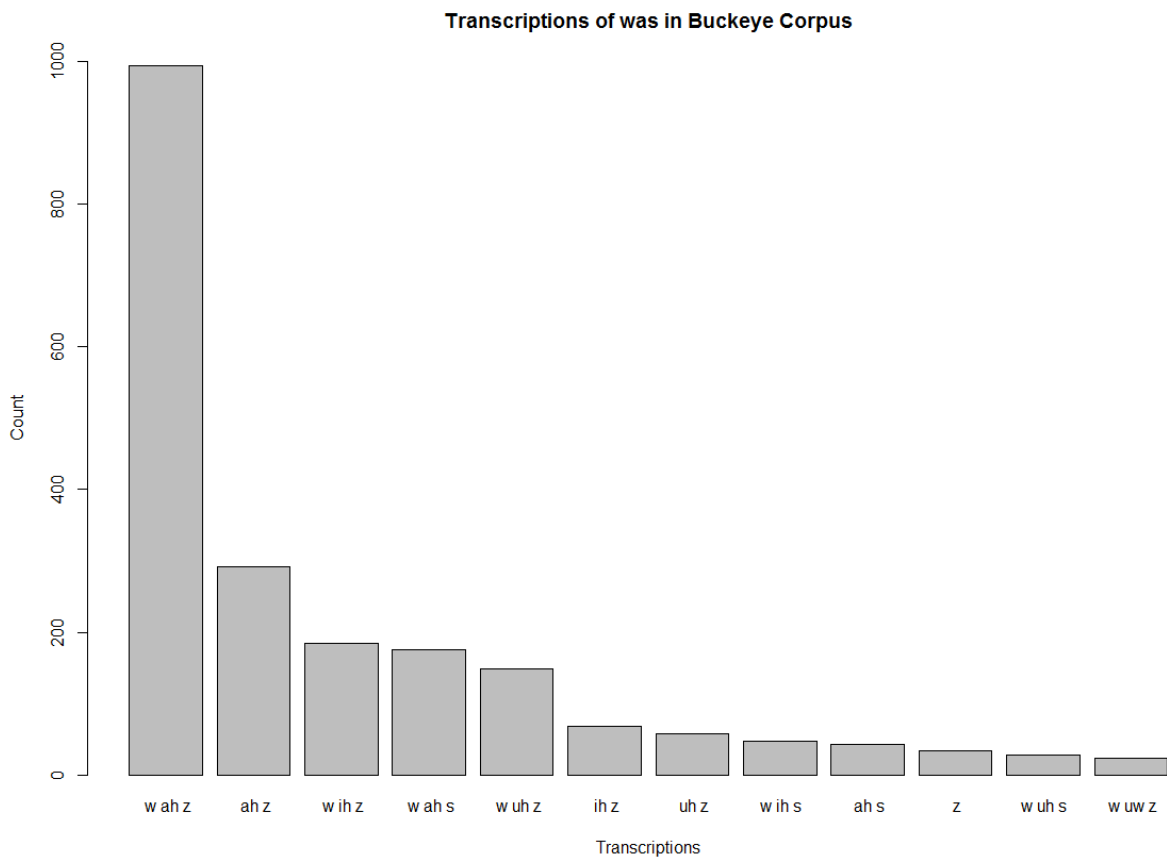


Figure 66. Distribution of *was* phonemic transcriptions in the Buckeye Corpus

In the Buckeye Corpus there are 25 different transcriptions of *is*. Figure 67 shows the distribution of transcriptions that occur 10 times or more. The citation form of [ih z] is by far the most common. Common reduced transcriptions show that often the final fricative is devoiced or the vowel is centralized. There are several instances where the transcription of full (non-contracted) *is* is [z] or [s], 35 and 17 times respectively. When listening to these specific instances in the corpus, it seems like an uncontracted variant was the speakers' intention, however there was simply a great deal of undershoot. These instances are not restricted to a handful of speakers, most speakers produced at least one of these 52 tokens. However, it may be the case that the *is* (as opposed to 's) of the transcriptions was somewhat influenced by knowledge of when contraction is likely or unlikely. Most of these instances are preceded by nouns, rather than pronouns (although there are a few cases of *there is* or *she is*) and many are preceded by words that preclude contraction such as *Texas is*, *else is*, *this is*, *place is*, *which is*, *yknow is*. Because it a rule of (written) Standard American English that contraction cannot happen with these words, the listener (transcriber) then hears *is* even though what is produced is [s] or [z].

In the Buckeye Corpus there are 25 different transcriptions of *are*. Figure 68 shows the distribution of transcriptions that occur 10 times or more. The citation form of [aa r] is not the most common. The most common form is the vowel [er] which in Buckeye's phonetic alphabet represents an unstressed rhotic vowel. Other forms have vowel centralization and/or lack the [r].

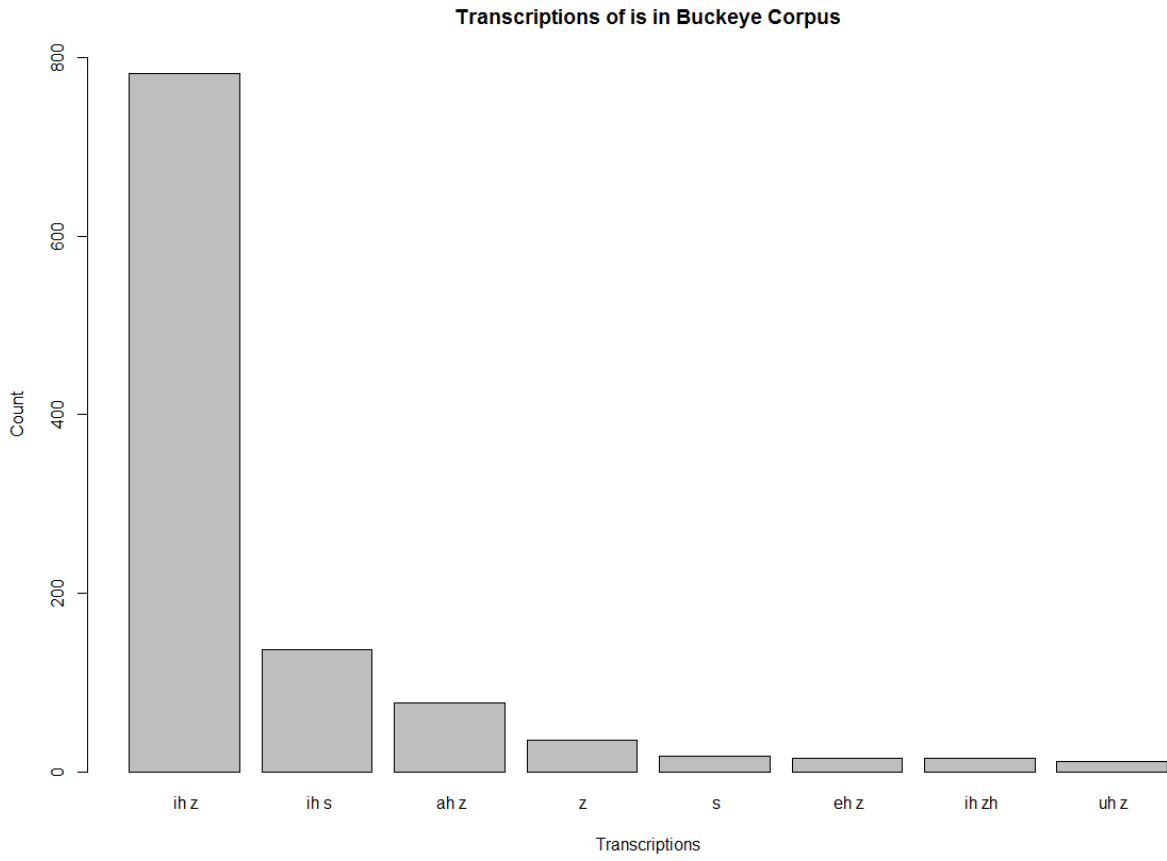


Figure 67. Distribution of *is* phonemic transcriptions in the Buckeye Corpus

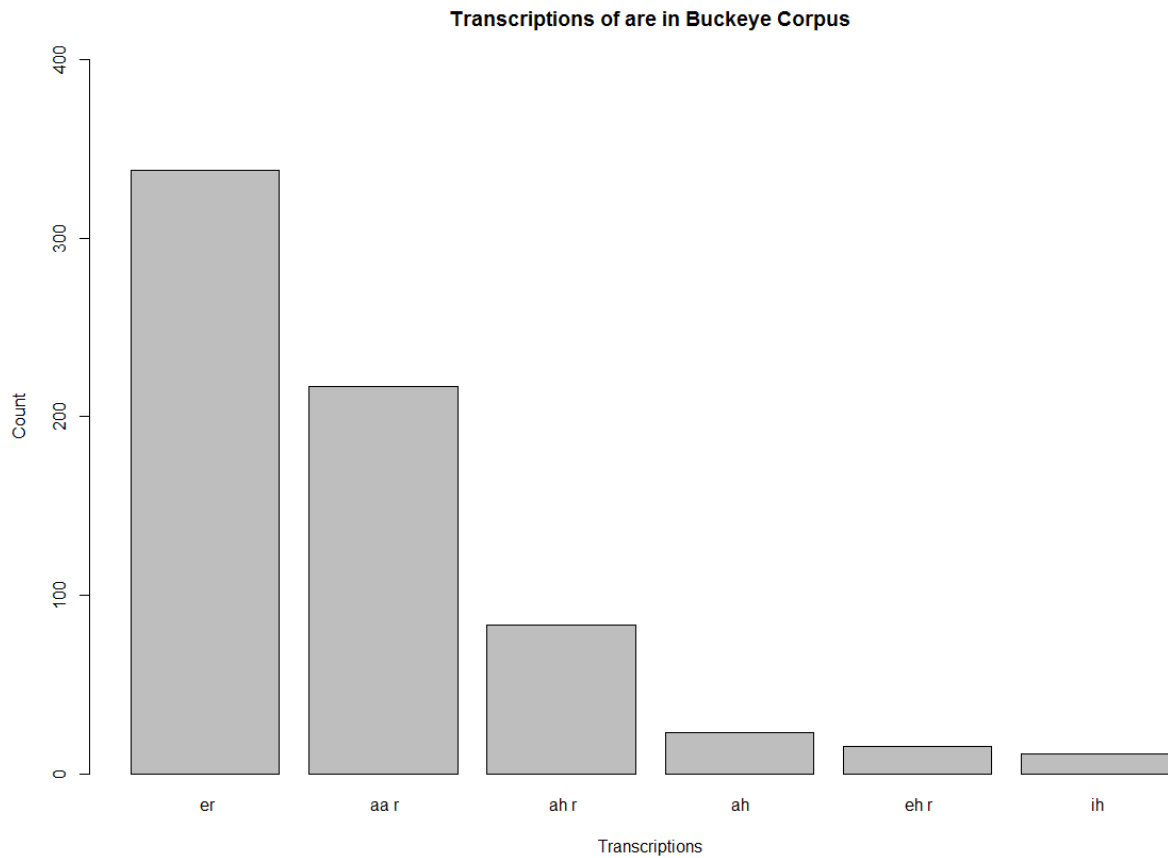


Figure 68. Distribution of *are* phonemic transcriptions in the Buckeye Corpus

In the Buckeye Corpus there are 34 different transcriptions of *were*, but most only occur once. Figure 69 shows the distribution of transcriptions that occur 5 times or more. The most common form is the glide [w] with the unstressed rhotic vowel [er]. The next most common form lacks the glide and only a few forms are transcribed with a full glide, vowel and approximate.

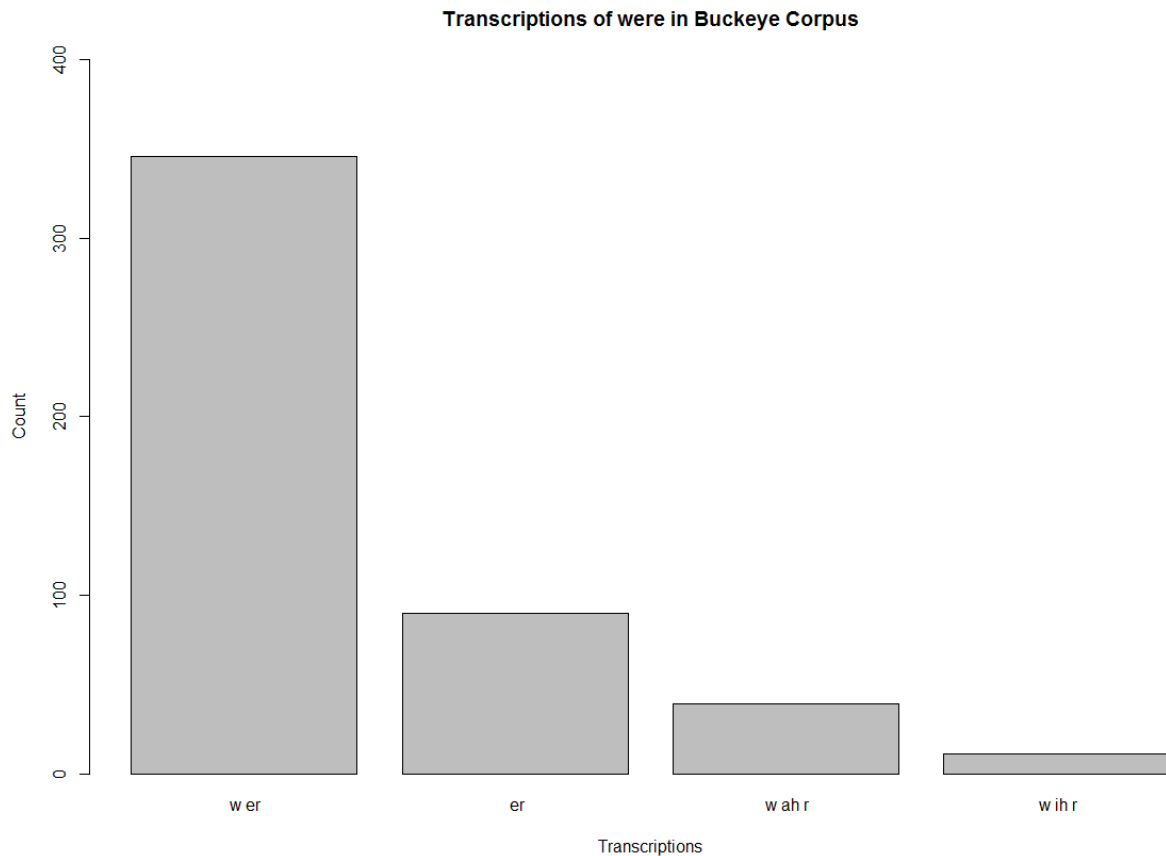


Figure 69. Distribution of *were* phonemic transcriptions in the Buckeye Corpus

Below, Table 22 presents the multimodel output for duration of *BE* words in the Buckeye Corpus. Fewer factors have high cumulative probability or significance than with the all words dataset. Construction, as with child and child-directed speech, had no probable effect on *BE* duration. The only significant variables were Following Transitional Probability, where high probability is associated with shorter targets, Preceding Contraction, where occurrence of contraction in the preceding ten words is associated with longer targets and Utterance Quartile, where targets in the final quartile of an utterance are longer than targets in preceding quartiles. The effects of these

variables are all in the same direction in the *BE* dataset as for the full dataset. Three more variables have high cumulative probability but did not reach significance. High Preceding Joint Probability is associated with longer targets (as with the full dataset), although neither Preceding Transitional Probability, nor an interaction between the preceding context variables had cumulative probability over 0.40. High Speaker Contraction Rate is associated with shorter targets and when a target follows a stressed syllable it tends to be shorter, as is the case for the all words dataset.

Table 22. IAS Duration Multimodel Inferencing Results: *BE*

Predictor	$\bar{\beta}$	$\bar{\sigma}$	LoCI	HiCI	<i>p</i>	CP
(Intercept)	0.774	0.128	0.523	1.024	0.000	NA
Following Transitional Probability	-0.100	0.008	-0.116	-0.084	0.000	1
Preceding Contraction - Y	0.059	0.011	0.038	0.079	0.000	1
Quartile:						
First (Reference Level)						0.97
Second	-0.003	0.011	-0.026	0.019	0.764	
Third	0.007	0.012	-0.017	0.031	0.543	
Fourth	0.053	0.019	0.016	0.089	0.005	
Preceding Joint Probability	0.005	0.005	-0.005	0.015	0.287	0.68
Speaker Contraction Rate	-0.159	0.148	-0.449	0.130	0.280	0.68
Preceding Stress Context - Y	-0.013	0.013	-0.040	0.013	0.323	0.65

Note: reported values are coefficients with shrinkage and adjusted standard error, $n = 4700$.

Table 23 presents the model comparison for the Buckeye *BE* dataset. Three models have a Δ below 2 and all perform much better than the null model. The best model includes all of the predictors, and the next best models include all predictors but Speaker Contraction Rate and Preceding Stress, respectively.

Table 23. Models of IAS duration with a Δ below 2

Model factors	k	df	log likelihood	AICc	Δ	w
123456	6	12	-1240.36	2504.79	0	0.32
23456	5	11	-1242.15	2506.36	1.57	0.15
12346	5	11	-1242.26	2506.57	1.78	0.13
(Null)	0	4	-1349.50	2707.01	202.22	0

Note: Cutoff: $\Delta < 2$, 1 = Speaker Contraction Rate, 2 = Following TP, 3 = Preceding Contraction, 4 = Preceding JP, 5 = Preceding Stress, 6 = Quartile

We now return to the research questions specified at the beginning of this section:

Q16. Are there construction specific effects for word shortening in the Buckeye Corpus data?

No. Just as with child speech and child-directed speech in the Redford Corpus, there were no word shortening effects of construction for *BE* in the Buckeye Corpus. Therefore we can conclude that the lack of construction-specific shortening in the Redford Corpus was not due to a low n , but because speakers do not utilize word length to signal construction meaning differences. In the Redford Corpus, contraction rates were significantly different for different construction meanings and so we can examine this effect with the larger Buckeye Corpus sample in section 4.3.2.3.

Q17. Are there word-specific effects for word shortening?

For one *BE* word (*am*) there were too few tokens to do a separate analysis. For all other *BE* words (*are, is, was, were*), there were no substantive differences between word-specific analyses and the full, collapsed analysis. For some of the words, fewer variables were significant, but for none of the words were additional variables significant. It does

not seem that any particular word is driving the effects shown in the full analysis. However, it seems that some words are more likely to undergo phonemic changes than others. For *is* and *was*, the citation forms were the most frequent found in the corpus, although there were many variants. For *are* and *were*, reduced (non-citation forms) were more frequent than citation forms. These words contain [r] and speakers were more likely to produce these function words with an unstressed vocalic [r] than a vowel and [r].

4.3.1.5. Multimodel Inferencing – Word Specific Results: HAVE

The primary research questions for this section are:

Q18. Are there construction specific effects for word shortening in the Buckeye Corpus data?

Q19. Are there word-specific effects for word shortening?

In order to answer these questions, I examine first the transcriptions of *HAVE* words as transcribed in the Buckeye Corpus and the distribution of those transcriptions. I then present a multimodel inferencing output for all *HAVE* instances in the corpus. No word-specific models were notably different from the collapsed *HAVE* model, and so no separate analyses are presented here.

In the Buckeye Corpus there are 40 different transcriptions of *had*, but most only occur once. Figure 70 shows the distribution of transcriptions that occur 5 times or more. The citation form [hh ae d] is the most common, but reduced commonly forms have a deleted initial, centralized vowel, or flap in place of the final stop or deleted final stops.

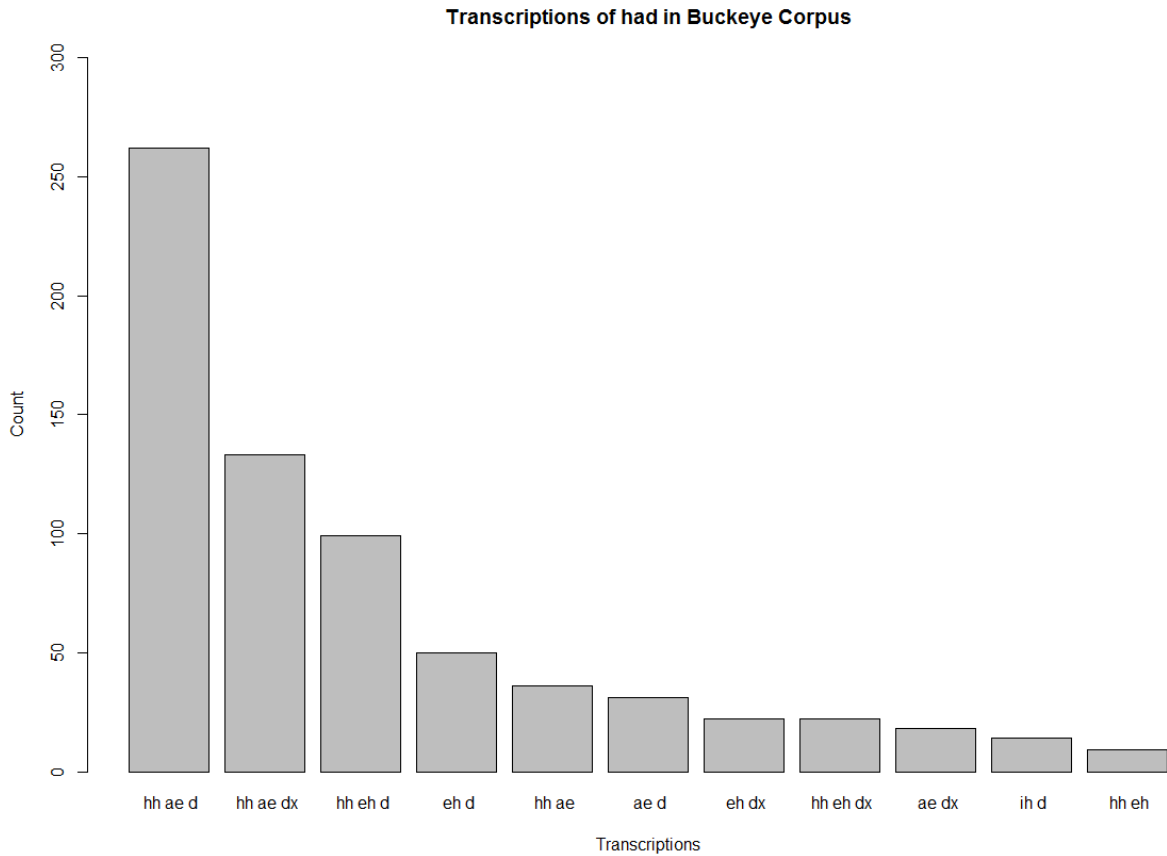


Figure 70. Distribution of *had* phonemic transcriptions in the Buckeye Corpus

In the Buckeye Corpus there are 29 different transcriptions of *has*, but most only occur once. Figure 71 shows the distribution of transcriptions that occur 5 times or more. The citation form [hh ae z] is the most common, but reduced commonly forms have a deleted initial, centralized vowel, or devoicing of the final fricative.

Transcriptions of has in Buckeye Corpus

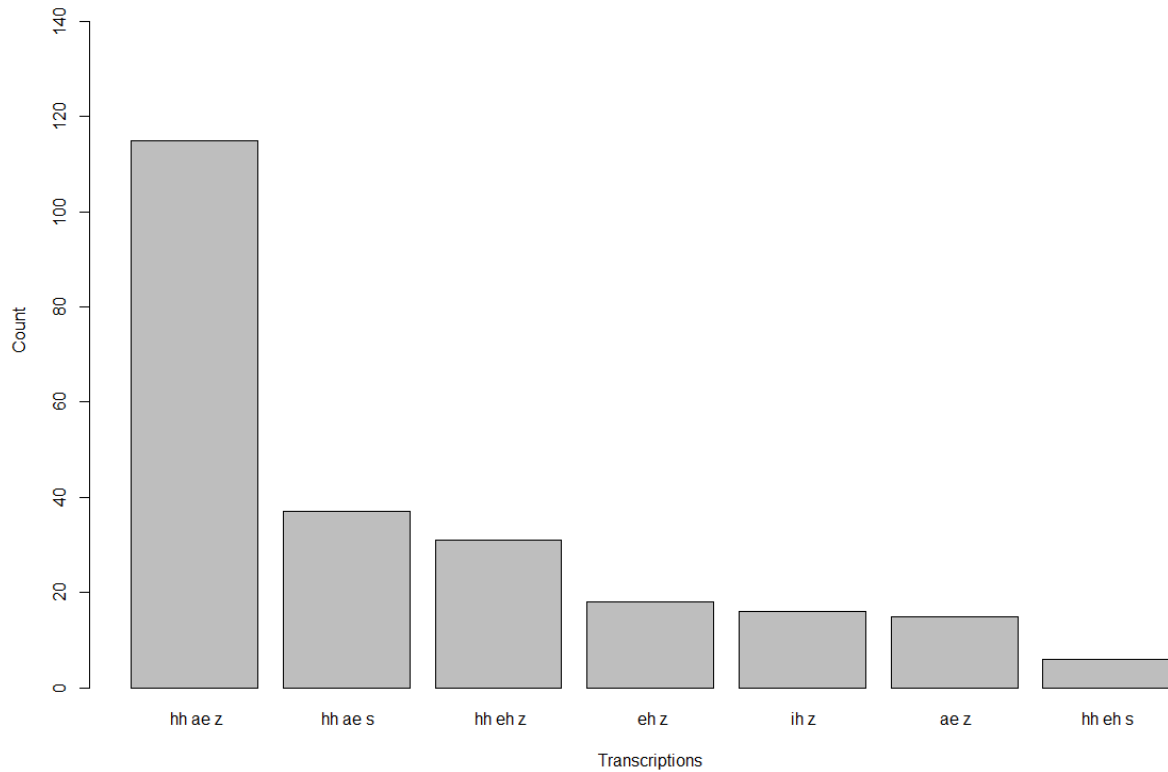


Figure 71. Distribution of *has* phonemic transcriptions in the Buckeye Corpus

In the Buckeye Corpus there are 70 different transcriptions of *have*, but most only occur once. Figure 72 shows the distribution of transcriptions that occur 10 times or more. The citation form [hh ae v] is the most common, but reduced commonly forms have a deleted initial, centralized vowel, or devoicing of the final fricative or deletion of the final fricative.

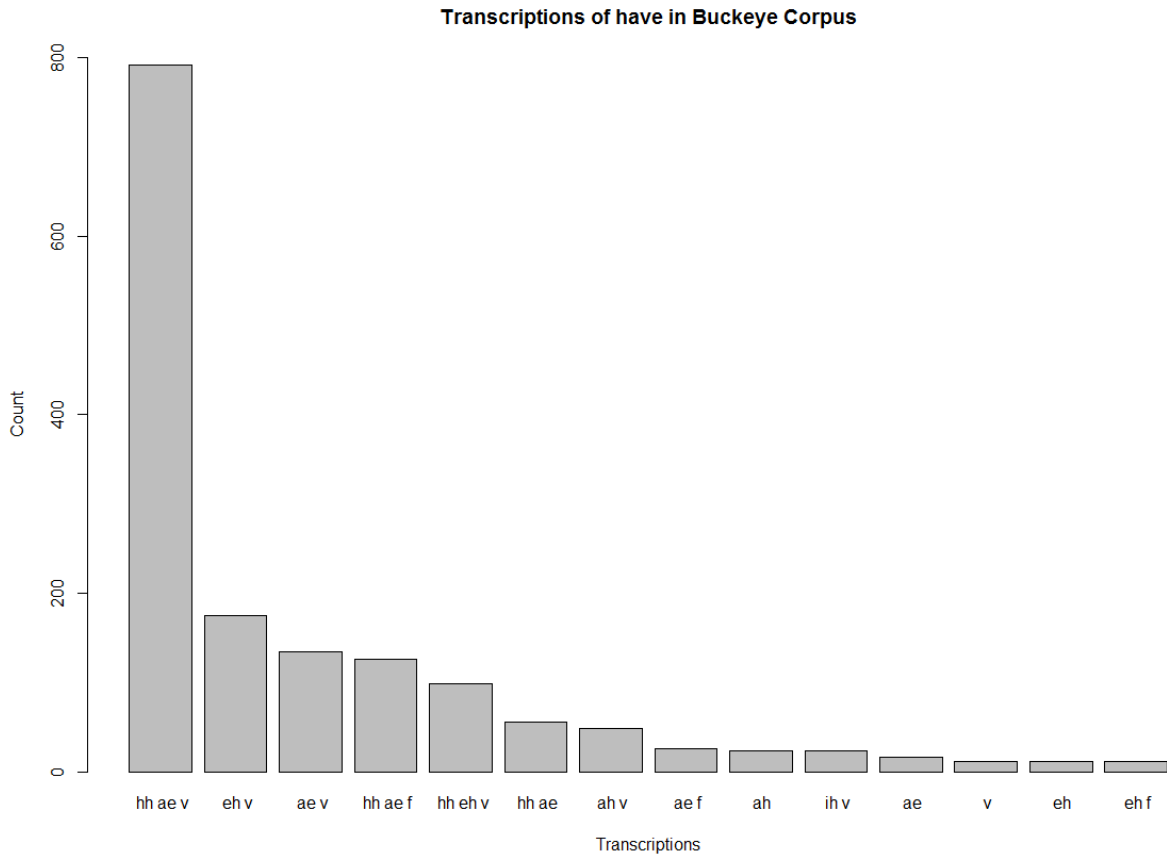


Figure 72. Distribution of *have* phonemic transcriptions in the Buckeye Corpus

Table 24 below shows the results for word shortening in the Buckeye Corpus for *HAVE*. There is a clear effect of construction, with perfect auxiliaries being shorter than possessive verbs and modal semi-auxiliaries, but possessive verbs and modal semi-auxiliaries showing no significant difference. As with *BE* tokens, high Following Transitional Probability or the presence of a stressed syllable preceding the target is associated with shorter targets. A target in the final utterance quartile is significantly lengthened compared to targets earlier in the utterance and high speaker contraction rate is associated with longer durations. Unlike with *BE* data or all data combined, preceding

contraction had no probable effect on duration. This indicates that either more data is necessary for this effect to show up as significant, or that *HAVE* tokens do not seem to be affected by preceding contraction, perhaps because they cannot contract, only shorten. In the Redford Corpus, *HAVE* words' duration was not affected by high speaker contraction rate, even though *BE* words were, in both child and child-directed speech.

Table 24. IAS Duration Multimodel Inferencing Results for *HAVE*

Predictor	$\bar{\beta}$	$\bar{\sigma}$	LoCI	HiCI	<i>p</i>	CP
(Intercept)	1.137	0.156	0.830	1.444	0.000	NA
Construction:						1
Possessive (reference level)						
Modal	-0.018	0.018	-0.054	0.018	0.339	
Perfect	-0.324	0.024	-0.371	-0.277	0.000	
Following Transitional Probability	-0.072	0.012	-0.096	-0.048	0.000	1
Preceding Stress Context - Y	-0.108	0.016	-0.139	-0.077	0.000	1
Quartile:						
First (Reference Level)						1
Second	0.017	0.017	-0.016	0.050	0.308	
Third	0.029	0.018	-0.006	0.065	0.107	
Fourth	0.117	0.024	0.070	0.164	0.000	
Speaker Contraction Rate	-0.142	0.195	-0.525	0.241	0.468	0.51

Note: reported values are coefficients with shrinkage and adjusted standard error, *n* = 2694.

Table 25 shows that the best two models have the full weight of all models combined. They differ only in their inclusion or exclusion of Speaker Contraction Rate as a factor. The other factors however, are highly likely to be influential in *HAVE* word shortening.

Table 25. Models of IAS duration with a Δ below 2 for *HAVE*

Model factors	k	df	log likelihood	AICc	Δ	w
12345	5	12	-946.96	1918.03	0	0.51
1345	4	11	-947.99	1918.09	0.06	0.49
(Null)	0	4	-1189.44	2386.90	468.87	0

Note: Cutoff: $\Delta < 2$, 1 = Construction, 2 = Speaker Contraction Rate, 3 = Following TP, 4 = Preceding Stress, 5 = Quartile

In the *HAVE* results for the Buckeye Corpus we saw results that paralleled those in the smaller Redford Corpus. We can now turn to the primary research questions for this section.

Q18. Are there construction specific effects for word shortening in the Buckeye Corpus data?

Yes. As with child and child-directed speech, perfect auxiliaries were significantly shorter than possessive, and in the Buckeye Corpus, perfect auxiliaries were also shorter than modal semi-auxiliaries. Additionally, as with the Redford Corpus data, Following Transitional Probability was significant, as was preceding stressed syllable. As with the child data, there was a significant difference between targets in the final quartile of an utterance and targets in the first three quartiles of an utterance.

Q19. Are there word-specific effects for word shortening?

No. *Have*, *has* and *had* all pattern similarly. Citation forms were always the most frequent form and no word-specific models showed any differences from the collapsed *HAVE* model. All *HAVE* words show the word shortening for perfect auxiliaries.

4.3.2. Contraction

4.3.2.1. *Word Distribution in Sample*

The above examination of word duration in the Buckeye Corpus confirmed results from the smaller Redford Corpus: Perfect auxiliaries are shorter than other variants of *HAVE*, high following transitional probability results in shorter words. However, the much larger n of data points allowed us to see some effects expected from the literature, but not fully realized in the smaller number of examples in the Redford Corpus, for instance, high joint probability can lead to longer words (Bell et al., 2003). In the examination of contraction below, there are much higher ns for the Buckeye Corpus, allowing us to confirm results from the Redford Corpus and also to examine perfect construction contraction. Table 26 below shows the distribution of contraction by lemma in the Buckeye Corpus.

In the Redford Corpus, there were too few instances of perfect constructions that were contracted to examine contraction in a model. In the Buckeye Corpus, there are overall more instances of the perfect construction, and therefore more instances of contraction, resulting in enough data to do a quantitative analysis. In Figure 73 below, we see rates of perfect construction contraction by speaker. In this figure, we see that all speakers vary in their production of perfect *HAVE*, producing both contracted and non-contracted variants. We also see that speakers contract more often than they produce full (non-contracted) forms.

Table 26. Contraction by Lemma in the Buckeye Corpus

Word	Contracted	Non-Contracted	Total
<i>am</i>	1008	32	1040
<i>are</i>	1478	502	1980
<i>had</i>	55	83	138
<i>has</i>	319	73	392
<i>have</i>	722	226	948
<i>is</i>	4,887	921	5,808
Total	8,469	1,837	10,306

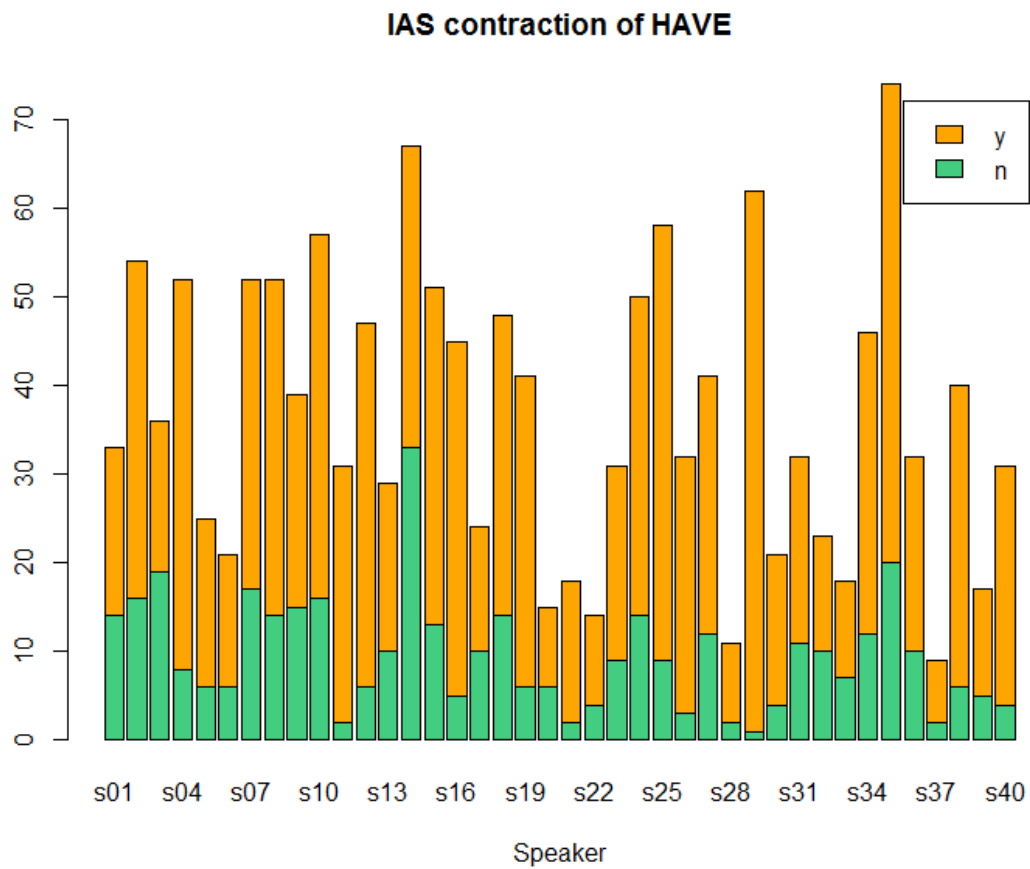


Figure 73. IAS Contraction rate of Perfect *HAVE* by Speaker

In addition to more examples of perfect constructions than in the Redford Corpus, there are many more examples of *BE* contraction in the Buckeye Corpus. Similar results across the analyses of the Redford Corpus and Buckeye Corpus would allow us to feel confident that the smaller dataset is showing a wider pattern. Figure 74 shows contraction rates of *BE* by speaker in the Buckeye Corpus. Like the adult and child speakers in the Redford Corpus, speakers in the Buckeye Corpus tend to contract more than they produce non-contracted forms. Because there is about ten times more contraction data in the Buckeye Corpus, we also see that all the speakers vary in their production of contracted and non-contracted variants. In the Redford Corpus, there were some instances of too few (one) variants per speaker and the speaker showed only contraction or non-contraction. In addition, a couple of the children produced 20~30 potentially contractible variants but contracted all of them.

4.3.2.2. *Random Forests*

In the random forests variable importance rankings below, the most important variables for contraction overall are clear. The top ranking variables were checked for collinearity for specific subsets before conduction multimodel inferencing. Variables that were collinear at $r = 0.5$ or higher were not included. In Figure 75, we see that the quartile that a word form appears in also has a important effect on its contraction. Figure 76 shows that more contraction occurs earlier in an utterance and proportionally more full forms occur as the utterance continues. However, the overall rate of contraction is so high that the lowest proportion of contracted forms is only about 50% in the fourth quartile.

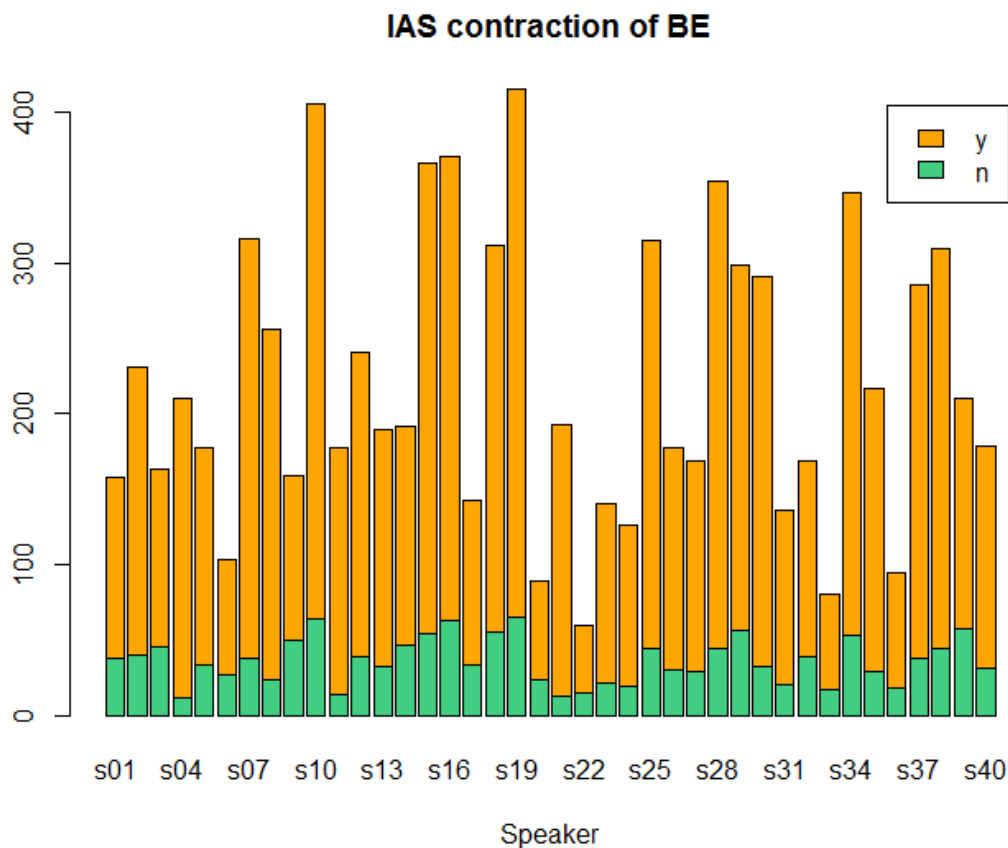


Figure 74. IAS Contraction rate of *BE* by Speaker

The variable Lemma is highly ranked, meaning that some *BE* and *HAVE* inflections (*am, are, had, has, have, is*) contract more often than others. As there are some independent variables that apply to only one word form or another, separate multimodel analyses are presented below for each word (*BE, HAVE*) and each inflection (*am, are, is, had, has, have*). Speaker is also highly ranked and will be a random effect in each of the regression models that follow. Speaker Contraction Rate is significant, and unlike for the duration random forests Speaker Contraction Rate that does not include *know* as a contraction is more predictive than Speaker Contraction Rate that does include *know* as a contraction. This obviously makes sense as the contraction DVs for

this analysis are only for *BE* and *HAVE*, and not for *yknow* or any other contractions. As the Contraction Rate IVs result in different levels of predictiveness, it seems that contraction of *yknow* is a different class of contraction. Speakers who contract *yknow* more (resulting in a different rate for *contr.rate* and *contr.rate.yknow*) do not necessarily contract *HAVE* or *BE* more. In the variable importance ranking, we also see that Local Syllable Rate is ranked higher than Average Syllable Rate, meaning the speed at which a person is talking makes more difference than their overall talking rate in predicting contraction. Finally, we see that age is ranked as important, but barely.

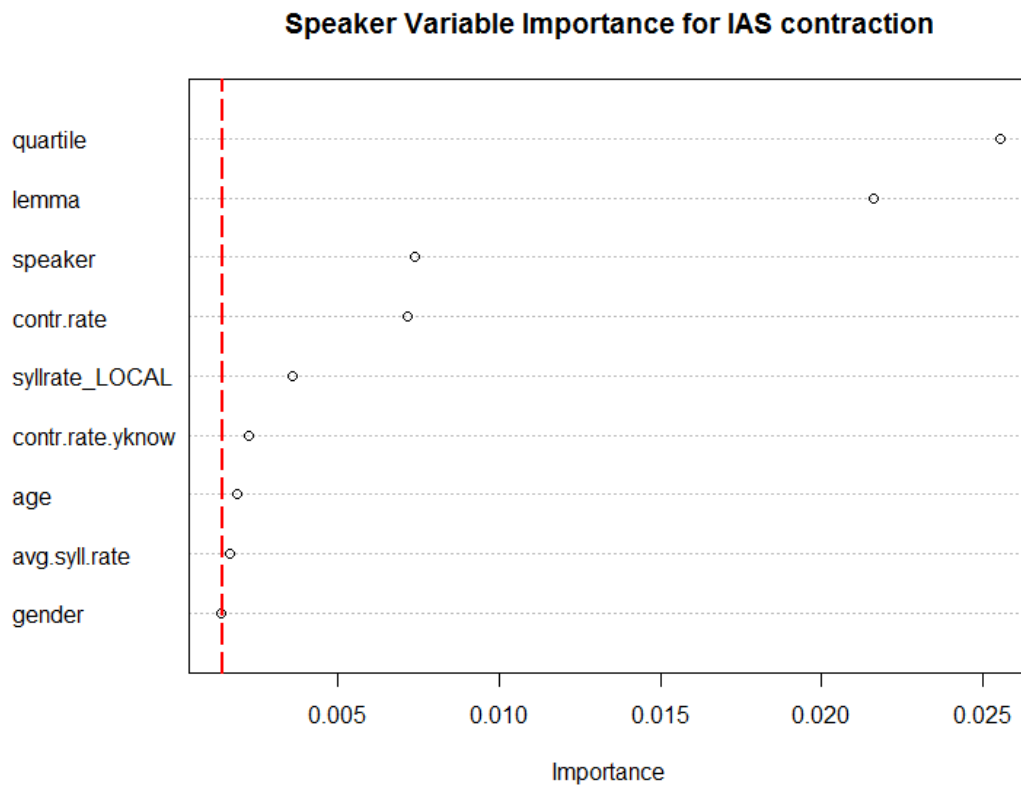


Figure 75. Random forest importance ranking of preceding context factors for contraction in IAS

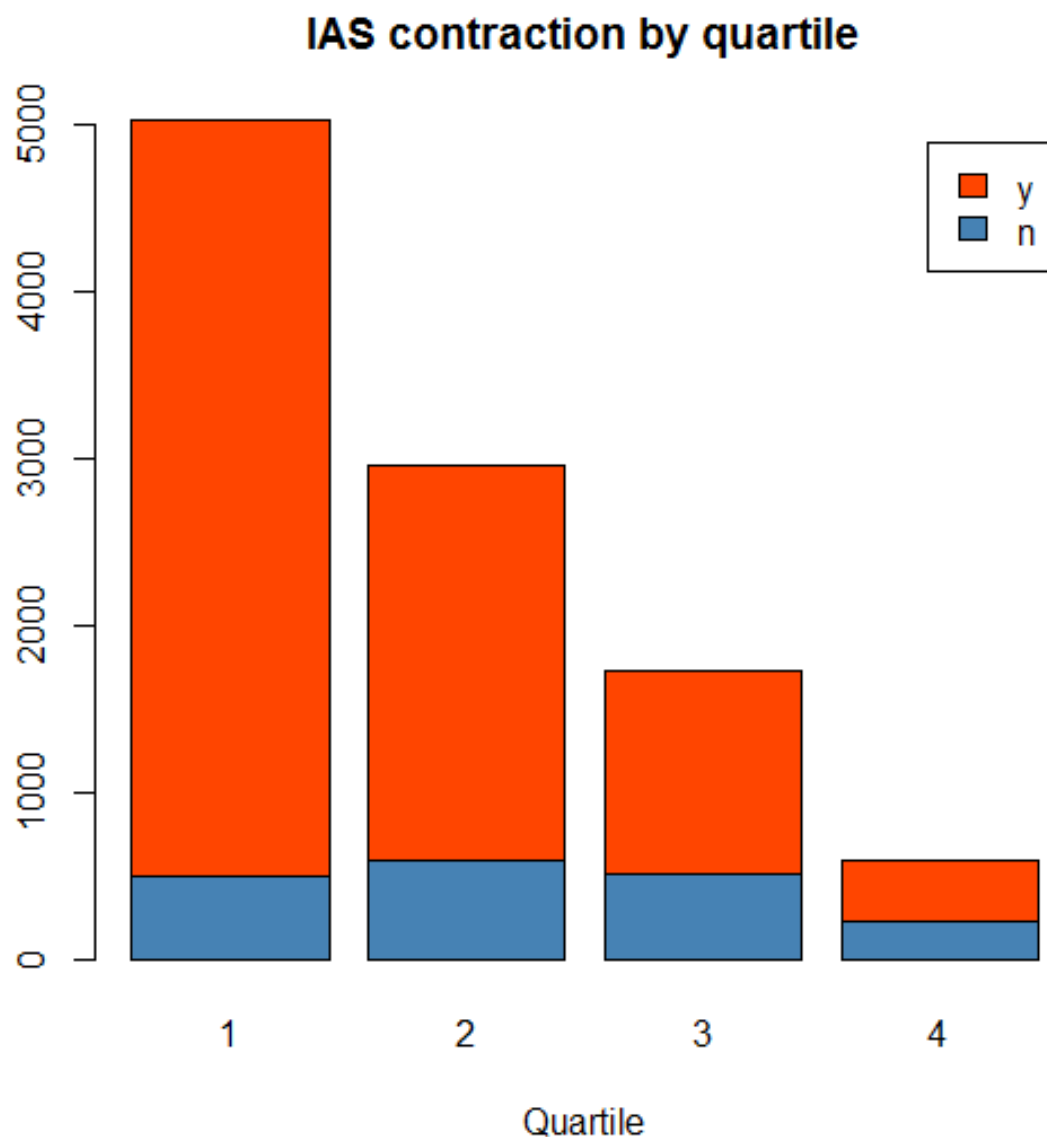


Figure 76. IAS Contraction by Utterance Quartile

Figure 77 below shows the importance of preceding context variables. Preceding contraction in the ten words preceding the target is important in predicting contraction of the target. Preceding Joint Probability and Preceding Noun Phrase Type are both ranked as important, with preceding JP being ranked as slightly more important (these variables are highly collinear for each of the data subsets). Bigrams with high joint probability in

the Buckeye Corpus include *it IS* (-2.606), *that IS* (-2.606), *I AM* (-2.963) *there IS* (-3.097), *you ARE* (-3.139), *he IS* (-3.141), *they ARE* (-3.143), *we ARE* (-3.180) and *I HAVE* (-3.209). Figure 78 below shows the distribution of contraction proportions for different NP types for *BE*. Clearly, targets following personal pronouns are much more likely to contract. The random forest variable importance ranking also shows that Preceding Transitional Probability is important. Bigrams with high transitional probability in the Buckeye Corpus include *there IS* (-0.234), *it IS* (-0.289), *they ARE* (-0.498) *we ARE* (-0.539), *might HAVE* (-0.548), *homosexuality IS* (-0.584), *must HAVE* (-0.584), *there ARE* (-0.596) and *what IS* (-0.612). Several of these bigrams have both high joint probability and transitional probability.

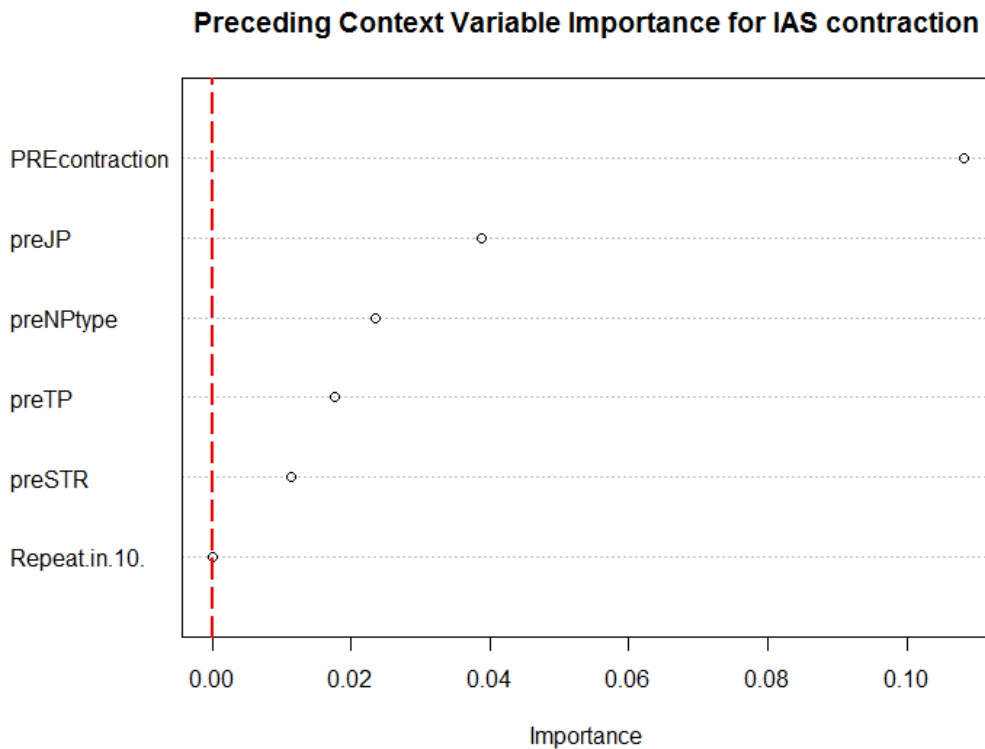


Figure 77. Random forest importance ranking of preceding context factors for contraction in IAS

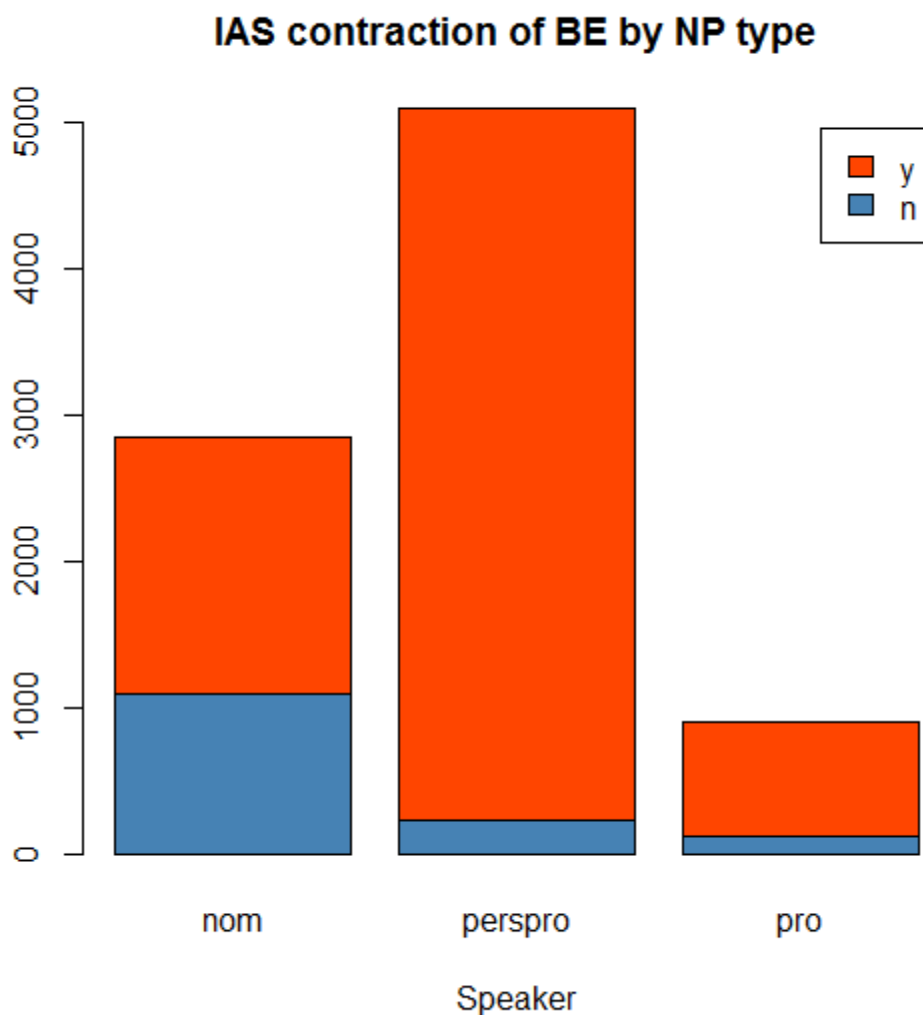


Figure 78. IAS Contraction Rate by NP Type for *BE*

Figure 79 below shows the variable importance ranking for the following context factors on contraction in the Buckeye corpus. Clearly Following Transitional Probability is the highest ranking factor, while Following Joint Probability, Following Stress and Construction Type are important, but less so. Examples of bigrams with high transitional probability include *IS happening* (-0.064), *AM sorry* (-0.093), *HAVE been* (-0.163) *HAVE gotten* (-0.201), *AM glad* (-0.240), *HAS been* (-0.335), *IS ridiculous* (-0.337), *HAVE seen* (-0.343) and *HAD grown* (-0.354).

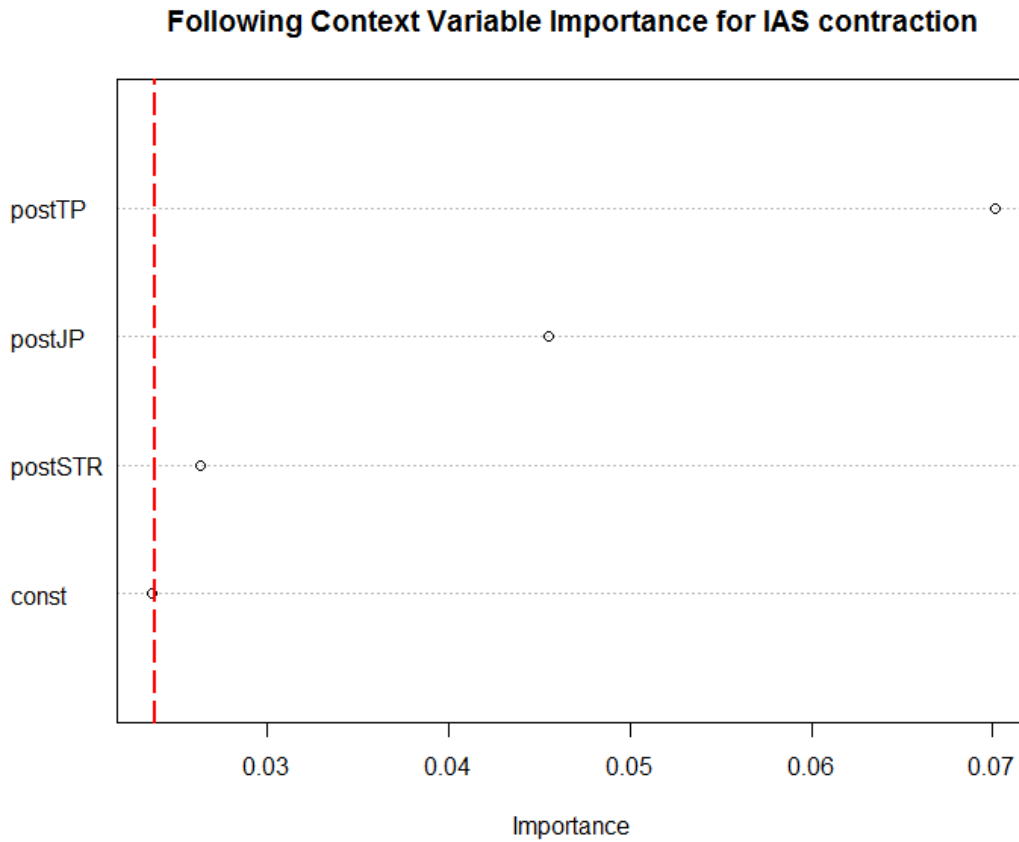


Figure 79. Random forest importance ranking of following context factors for contraction in IAS

4.3.2.3. Multimodel Inferencing – Word Specific Results: BE

4.3.2.3.1. All BE Data Collapsed

This section examines the influences on contraction for inflections of *BE*: *is*, *am* and *are*. The main research questions for this section are:

Q20. Is *BE* contraction influenced by the same factors as *BE* contraction in the Redford Corpus?

Q21. Because there are enough tokens of each inflection of *BE* (*is*, *are*, *am*), we can investigate particularities of the individual word forms. Are there word-specific characteristics for contraction?

Table 27 below presents the multimodel output for contraction of *BE* (that is, *am*, *are*, and *is*). Positive values of coefficients are associated with more contraction, negative values are associated with less contraction. Several factors have a cumulative probability of one, meaning they are extremely likely to influence contraction. First, we see that contraction begets contraction. A speaker with a high overall rate of contraction is highly likely to contract in a given instance, if a speaker has contracted *HAVE* or *BE* in the previous ten words, they are likely to contract again. The inflections *am*, *are*, and *is* are associated with different contraction rates. *Am* is extremely likely to contract, *is* less so and *are* even less so. In model testing when *is* was used as the reference level for this variable, *are* contracted significantly less than *is*. Construction Type is also highly predictive of contraction with both the future and progressive constructions being associated with higher rates of contraction than the copula construction. The passive construction does not differ significantly than the copula construction. High following transitional probability and high preceding joint probability are also associated with high levels of contraction, meaning the more probable the context, the more likely contraction is to occur. The final variable with a cumulative probability of one is Utterance Quartile.

The highest rates of contraction are likely to occur in the first quartile of an utterance. Each following quartile is less likely to have contracted forms than the quartile preceding. Additionally, as seen in Figure 76 above, *BE* and *HAVE* are much more likely to occur in the first quartile of an utterance than each subsequent quartile. Speaker Syllable Rate also has a very high cumulative probability (0.99), and before shrinkage this factor had a positive coefficient showing that higher syllable rates are associated with more contraction, so the slower a speaker produced an utterance, the more likely they were to contract. The last factor that has some predictiveness is the following stress context. When the syllable following the target is stressed, the target is more likely to be contracted than when the following syllable is unstressed. Factors that did not achieve high cumulative probability, and are therefore unlikely to be predictive for contraction are Following Joint Probability and Speaker Age.

Model rankings for *BE* contraction are presented Table 28 below. There is one clear best model accounting for most of the model weight. The next best model has a delta below 2 (2.15) and accounts for most of the remaining weight ($w = 0.23$). All of the factors included in the model are important and there is no model using this set of predictors that comes in a close second. In this case, a traditional step-wise model selection approach would come up with the same model ranked highest in the multimodel output procedure (but in essence assign it 100% probability).

Table 27. IAS Contraction of *BE* Multimodel Inferencing Results

Predictor	$\bar{\beta}$	$\bar{\sigma}$	LoCI	HiCI	<i>p</i>	CP
(Intercept)	1.582	0.049	1.485	1.679	0.000	NA
Speaker Contraction Rate	0.306	0.063	0.182	0.429	0.000	1
Preceding Contraction - Y	0.618	0.009	0.600	0.637	0.000	1
Word Form:						
<i>AM</i> (Reference Level)						1
<i>ARE</i>	-0.108	0.009	-0.126	-0.090	0.000	
<i>IS</i>	-0.037	0.009	-0.055	-0.019	0.000	
Construction:						
Copula (reference level)						1
Future	0.056	0.013	0.031	0.080	0.000	
Passive	0.029	0.020	-0.011	0.069	0.151	
Progressive	0.054	0.008	0.038	0.070	0.000	
Following Transitional Probability	0.035	0.005	0.025	0.045	0.000	1
Preceding Joint Probability	0.147	0.003	0.141	0.152	0.000	1
Quartile:						
First (Reference Level)						1
Second	-0.016	0.006	-0.028	-0.005	0.005	
Third	-0.037	0.007	-0.051	-0.023	0.000	
Fourth	-0.097	0.011	-0.119	-0.075	0.000	
Speaker Syllable Rate	0.000	0.000	0.000	0.000	0.001	0.99
Following Stress Context - Y	0.007	0.008	-0.009	0.023	0.377	0.84

Note: reported values are coefficients with shrinkage and adjusted standard error, $n = 8401$.

Table 28. Models of IAS contraction of *BE* with a Δ below 2

Model factors	<i>k</i>	<i>df</i>	log likelihood	AICc	Δ	<i>w</i>
123456789	9	17	836.34	-1638.61	0	0.69
(Null)	0	3	-3731.3	7468.61	9107.22	0

Note: Cutoff: $\Delta < 2$, 1 = Construction, 2 = Speaker Contraction Rate, 3 = Lemma, 4 = Following Stress, 5 = Following TP, 6 = Preceding Contraction, 7 = Preceding JP, 8 = Quartile, 9 = Speaker Syllable Rate

4.3.2.3.2. *is-'s*

Table 29 shows the multimodel inferencing results for contraction of *is* in the Buckeye Corpus. As with *BE* contraction, contraction of one of the ten preceding words results in a higher likelihood of contraction of the target. When the subject noun phrase is a non-personal pronoun (*there, what, who, everyone, everybody, here, nobody, nowhere, etc.*), there is not a significant difference in rates of contraction than when the subject NP is a nominal. Targets following subject NPs that are personal pronouns (*he, it, she*), are more likely to contract. Targets in highly probable contexts (with Following Transitional Probability and Preceding Joint Probability) are more likely to contract. However, targets with high Following Joint Probability are actually significantly less likely to contract (cumulative probability = 0.92). Targets in the first quartile of an utterance are more likely to contract than in any later quartile. Targets in the final quartile of an utterance are the least likely to contract. As there are significantly more targets in the first quartile, decreasing in each subsequent quartile, this also means that targets in the most probable context, as measured by quartile, are also the targets most likely to contract. This is an additional kind of context probability beyond bigram probability that affects contraction. Speakers who have high contraction rates generally are significantly more likely to contract. Construction Type also has a high cumulative probability (0.99), with the

progressive construction being associated with much higher rates of contraction than the copula construction. The passive and future constructions do not have significantly different rates of contraction from the copula construction. Targets followed by a stressed syllable are more likely to contract than targets followed by an unstressed syllable. Finally, the last factor with a cumulative probability over 0.5 is Syllable Rate, with utterances produced at a higher (slower) rate associated with more contraction.

The model rankings for *IS* contraction are presented in Table 30 below. The best performing model has a very high weight of 0.56. This model includes all factors. The next best model includes all factors except for syllable rate. Both of these models reasonably account for the data.

4.3.2.3.3. *are- 're*

Although there were some instances of *are* contracting after a non-personal pronoun target, such as *there 're* and *where 're*, these were infrequent, so *are* contraction was only examined in the context of personal pronoun + *ARE*. This meant that there were only a few different contexts where *are* contraction could be examined ($n = 1554$). The preceding bigram joint probabilities for the *are- 're* data subset are *you ARE* (-3.139), *they ARE* (-3.143) and *we ARE* (-3.180). This was not significant when examined with multimodel inferencing. Only a few factors showed high cumulative probability and significance: Speaker Contraction Rate, Preceding Contraction and Following Transitional Probability. These effects were all in the same direction as the collapsed *BE* analysis, so no further results for *ARE* contraction are presented here.

Table 29. IAS Contraction of *IS* Multimodel Inferencing Results

Predictor	β^-	σ^-	LoCI	HiCI	<i>p</i>	Cumul Prob
(Intercept)	1.440	0.059	1.324	1.556	0.000	NA
Preceding Contraction – Y	0.662	0.012	0.638	0.685	0.000	1
Following Transitional Probability	0.050	0.007	0.036	0.065	0.000	1
Preceding Joint Probability	0.121	0.004	0.114	0.128	0.000	1
Quartile:						
First (Reference Level)						1
Second	-0.019	0.007	-0.033	-0.006	0.005	
Third	-0.064	0.009	-0.081	-0.047	0.000	
Fourth	-0.163	0.014	-0.190	-0.135	0.000	
Speaker Contraction Rate	0.302	0.074	0.157	0.447	0.000	1
Construction:						0.99
Copula (reference level)						
Future	0.027	0.019	-0.010	0.064	0.155	
Passive	-0.020	0.027	-0.072	0.032	0.457	
Progressive	0.049	0.013	0.023	0.075	0.000	
Preceding NP Type:						
Nominal (Reference Level)						0.98
Personal Pronoun	-0.015	0.008	-0.031	0.001	0.061	
Non-Personal Pronoun	0.013	0.009	-0.005	0.032	0.155	
Following Joint Probability	-0.008	0.004	-0.016	0.000	0.043	0.92
Following Stress Context – Y	0.017	0.010	-0.001	0.036	0.071	0.89
Speaker Syllable Rate	0.000	0.000	0.000	0.000	0.255	0.71

Note: reported values are coefficients with shrinkage and adjusted standard error, *n* = 5807.

Table 30. Models of IAS contraction of *IS* with a Δ below 2

Model factors	k	df	log likelihood	AICc	Δ	w
1/2/3/4/5/6/7/8/9/10	10	18	534.2	-1032.27	0	0.56
1/2/3/4/5/6/7/8/9	9	17	532.33	-1030.56	1.72	0.24
(Null)	0	3	-2358.63	4723.27	5755.54	0

Note: Cutoff: $\Delta < 2$, 1 = Construction, 2 = Speaker Contraction Rate, 3 = Following JP, 4 = Following Stress, 5 = Following TP, 6 = Preceding Contraction, 7 = Preceding JP, 8 = Preceding NP Type, 9 = Quartile, 10 = Speaker Syllable Rate

4.3.2.3.4. *am-'m*

The contraction of *AM* is an interesting case because the overall rate of contraction is so high (96.82% of the time *am* is contracted in the Buckeye Corpus). However, as seen in Table 31, there are several factors which are associated with particularly high contraction, including Preceding Contraction, high Following Transitional Probability and Quartile.

Table 31. IAS Contraction of *AM* Multimodel Inferencing Results

Predictor	β	σ	LoCI	HiCI	p	CP
(Intercept)	1.044	0.038	0.970	1.118	0.000	NA
Preceding Contraction - Y	0.970	0.031	0.910	1.030	0.000	1
Following Transitional Probability	0.023	0.006	0.011	0.035	0.000	1
Quartile:						
First (Reference Level)						0.99
Second	-0.033	0.009	-0.052	0.015	0.000	
Third	0.001	0.011	-0.021	0.023	0.920	
Fourth	0.014	0.019	-0.024	0.052	0.477	
Following Joint Probability	-0.003	0.004	-0.012	0.005	0.442	0.53

Note: reported values are coefficients with shrinkage and adjusted standard error, $n = 1040$.

Table 32 shows that the best ranked model includes all of the factors. The next best model includes all factors but one: Following Joint Probability. The factors contained in both models, Following Transitional Probability, Preceding Contraction and Quartile are highly likely to be influential in *AM* contraction.

Table 32. Models of IAS contraction of *AM* with a Δ below 2

Model factors	k	df	log likelihood	AICc	Δ	w
1234	4	9	720.51	-1422.85	0	0.53
234	3	8	719.36	-1422.58	0.26	0.46
(Null)	0	3	350.8	-695.58	727	0

Note: Cutoff: $\Delta < 2$, 1 = Following JP, 2 = Following TP, 3 = Preceding Contraction, 4 = Quartile

In Figure 80 we see that normally *AM* shows up in the first two quartiles of the utterance. There are examples of *AM* in the last half of the utterance but these are much rarer: “*out in the real world I’m talking about*” or “*if they open up the intern program like I’m hoping they will*”. In any case, there is only a significant difference in *AM* contraction between the first two quartiles, with contraction more likely in the first quartile of the utterance. Bigrams with high following transitional probability include *AM sorry* (-0.093), *AM glad* (-0.240), *AM guessing* (-0.306), *AM afraid* (-0.459), *AM amazed* (-0.672), and *AM sure* (-0.687). These top probable combinations reflect expressions of subjectivity (“expression of self and the representations of a speaker’s (or, more generally, a locutionary agent’s) point of view in the discourse” [Finegan 1995:1]), where speakers are generally commenting on their feelings about other content in the utterance: “*I’m sorry let him pay the penalty*” or “*yknow what I mean they shouldn’t even be out on the street I’m sorry,*” although there are examples that do not reflect subjectivity: “*I’m*

afraid sometimes that I'll die before I get a chance to do." Other factors do not reach a high level of cumulative probability for AM contraction.

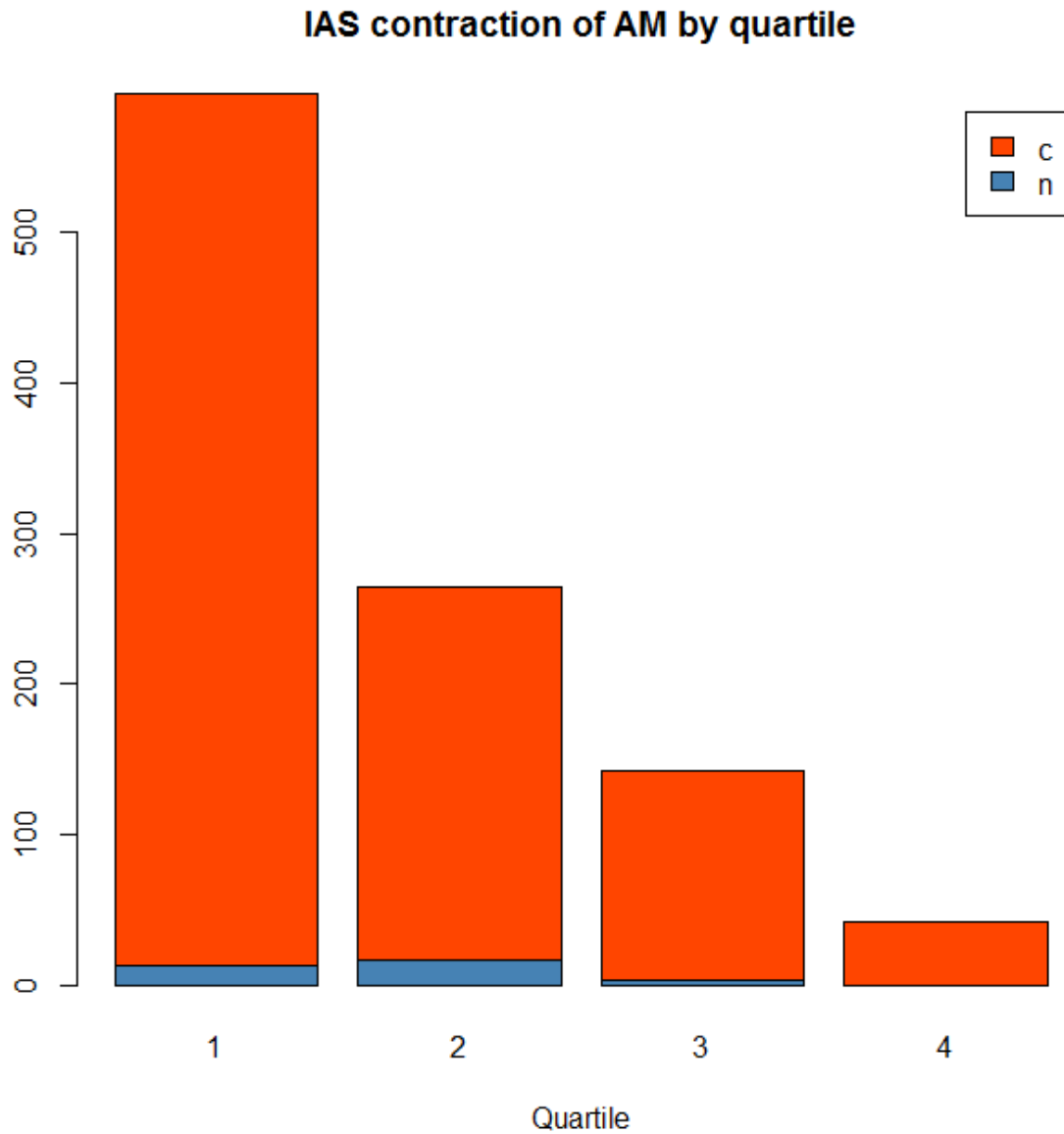


Figure 80. IAS Contraction of AM by Utterance Quartile

4.3.2.3.5. Conclusion

We now return to the research questions put forth at the beginning of this section.

Q20. Is *BE* contraction influenced by the same factors as *BE* contraction in the Redford Corpus?

Yes. In both the Redford Corpus and Buckeye Corpus, high preceding probability context is associated with more contraction, as are preceding instances of contraction, future and progressive constructions, position in the first utterance quartile and high overall speaker contraction rate. Only in the Buckeye Corpus data is Speaker Syllable Rate an important variable for contraction, with slower speakers contracting more often than faster ones.

Q21. Are there word-specific characteristics for contraction?

Yes. *IS* contracts with more kinds of subject NPs, but still contracts particularly often with personal pronouns. *AM* contracts at very high rates, and a closer look at the data showed that it contracts often in cases of speaker subjectivity.

4.3.2.4. Multimodel Inferencing – Word Specific Results: HAVE

4.3.2.4.1. All *HAVE* Data Collapsed

In Chapter III, it was not possible to investigate the factors influencing contraction of the perfect auxiliary, due to low occurrence of this construction, particularly for contracted auxiliaries and particularly in child speech. There are, however, many instances of the perfect auxiliary, contracted and non-contracted, in the Buckeye Corpus. The main research questions for this section are:

Q22. Is contraction of *HAVE* influenced by the same factors that influence the word shortening of *HAVE*?

Q23. Is contraction of *HAVE* influenced by the same factors that influence the contraction of *BE*?

Q24. Because there are enough tokens of each inflection of *HAVE* (*have, has, had*), we can investigate particularities of the individual word forms. Are there word-specific characteristics for contraction?

Figure 81 shows the distribution of contracted and non-contracted *HAVE* perfect auxiliaries by word form. As seen in Figure 81, there are many more instances of *have-’ve* in the corpus than other inflections of *HAVE*, however it is *HAS* that contracts proportionally most often. In Figure 82, we see that auxiliaries contract often with *have~’ve* as well, as in *must’ve, should’ve, might’ve*, etc. Some nominal subjects contract with *have* and *has* perfect auxiliaries as well: “all their life *people’ve been saying*, “*all the work Gore’s been doing*,” and “*because that’s what Dad’s gotta do*.” Excluded from analyses are instances of *HAVE* that could not contract, including *had* with nominal subject NPs and cases where *has* follows a word ending in a sibilant.

There are a few instances of *have* contracting with nominal subjects, as indicated in Figure 82. There are three instances of *people’ve*. There are only a few examples of *have* contracting with a non-personal pronoun. These include *who’ve, there’ve* and *that’ve*.

IAS contraction of HAVE by lemma

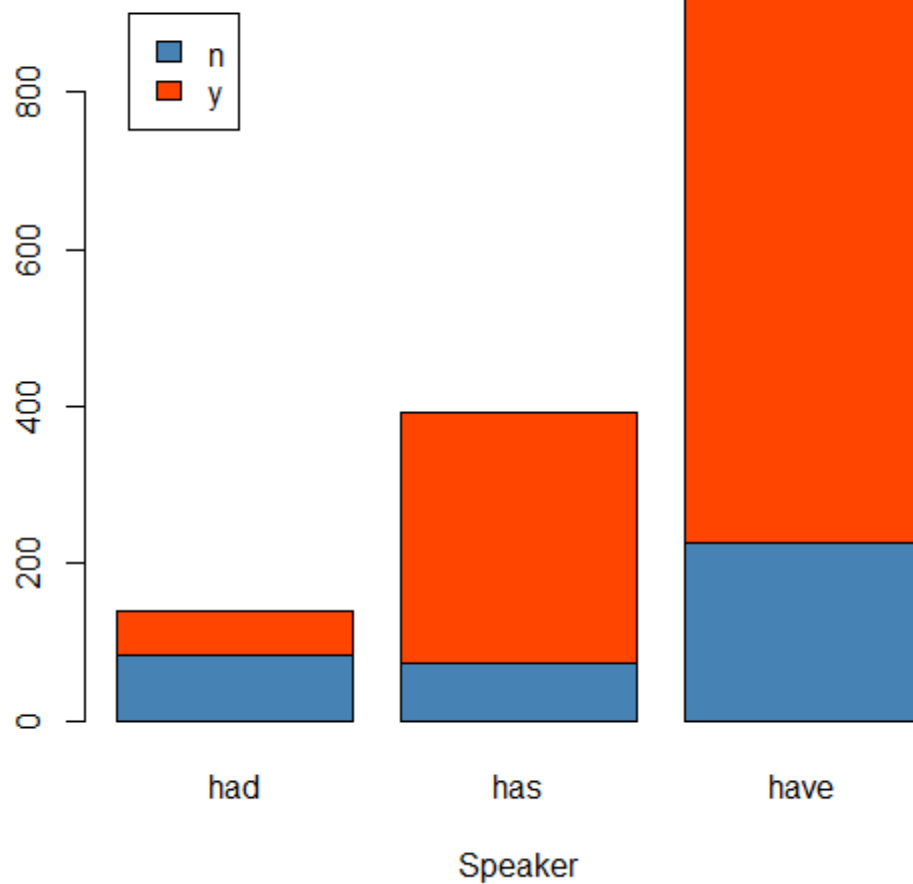


Figure 81. IAS Contraction of *HAVE* by Inflection

IAS contraction of have~'ve by NP type

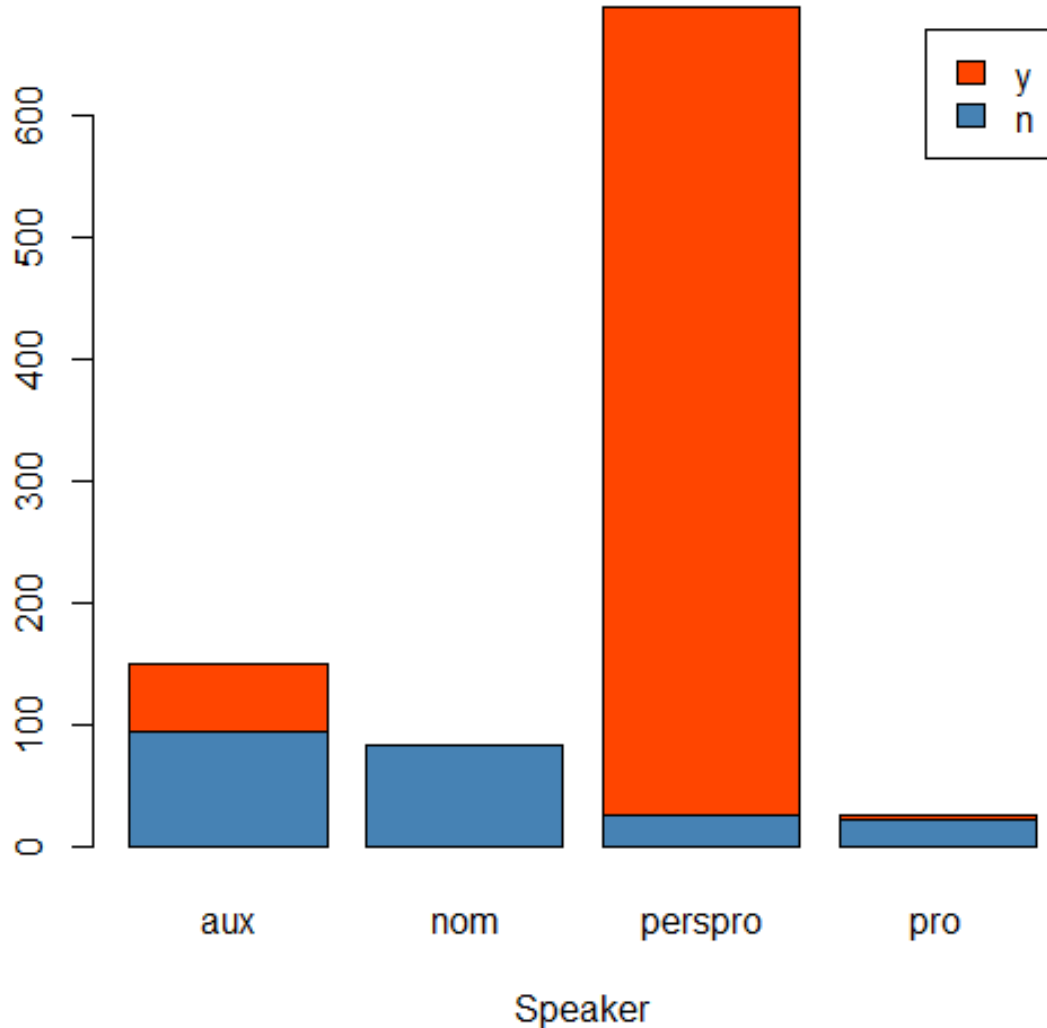


Figure 82. IAS Contraction of *have~'ve* by Subject Noun Phrase Type

Table 33 shows that Preceding Contraction (of any word) occurring before a target makes it more likely to contract. However, Speaker Contraction Rate has no probable effect on contraction of perfect auxiliaries, even though it did on perfect auxiliary reduction in child and child-directed speech. The inflection *has-'s* is the

inflection most likely to contract. High Preceding Joint Probability and Preceding Transitional Probability increase the likelihood of contraction. Speakers with higher syllable rates (slower speakers) are more likely to contract, although age and gender has no effect on perfect auxiliary contraction. Perfect auxiliaries that are followed by a stressed syllable (*he'd seen, he's tried*) are more likely to contract than perfect auxiliaries that are followed by an unstressed syllable (*he had developed, he has become*). Finally, perfect auxiliaries are significantly most likely to contract when they are in the first quartile of the utterance than any other quartile, although there is no significant difference between any of the other following quartiles.

There is one clear best performing model for this dataset, and it contains all of the factors included. The next best model (not shown in Table 34 below) has a delta of 8.15, indicating that it is much worse performing than the best model.

4.3.2.4.2. *have- 've*

There are several potential variables to investigate for preceding context: Preceding Joint Probability, Preceding Transitional Probability and Preceding NP Type. Preceding NP Type is not correlated strongly with either of the probability variables, but the latter two are correlated with each other. Preceding Transitional Probability performs better than Preceding Joint Probability for predicting *HAVE* contraction, so it was included in the multimodel analysis. As seen in Table 35, perfect auxiliaries followed by personal pronouns are more likely to contract than auxiliaries followed by any other subject type. Auxiliaries in high preceding probability contexts are more likely to contract, but those in low following probability contexts are actually more likely to contract, making the coefficient for Post Transitional Probability negative. Despite this,

the bigram with highest following transitional probability, *HAVE been*, is more likely to contract than not (63% of the time). As with the collapsed *HAVE* data, slower speakers contract more often and if a speaker has contracted recently, they are likely to contract again.

There is one model that performs well, and it is the model including all of the factors. The next best model has a delta of 2.13 and lacks Speaker Syllable Rate, as seen in Table 36.

Table 33. IAS Contraction of *HAVE*, *HAD* and *HAS* Multimodel Inferencing Results

Predictor	β	σ	LoCI	HiCI	p	CP
(Intercept)	1.441	0.064	1.316	1.566	0.000	NA
Preceding Contraction - Y	0.728	0.022	0.685	0.772	0.000	1
Word Form:						
<i>HAD</i> (Reference Level)						1
<i>HAS</i>	0.198	0.029	0.141	0.256	0.000	
<i>HAVE</i>	0.023	0.027	-0.030	0.076	0.393	
Preceding Joint Probability	0.101	0.013	0.076	0.126	0.000	1
Quartile:						
First (Reference Level)						1
Second	-0.064	0.017	-0.097	-0.030	0.000	
Third	-0.071	0.021	-0.112	-0.031	0.001	
Fourth	-0.062	0.029	-0.118	-0.006	0.029	
Speaker Syllable Rate	0.001	0.000	0.000	0.001	0.001	0.99
Preceding Transitional Probability	0.090	0.028	0.035	0.145	0.001	0.99
Following Stress Context - Y	0.051	0.017	0.017	0.084	0.003	0.98

Note: reported values are coefficients with shrinkage and adjusted standard error, $n = 1478$.

Table 34. Models of IAS contraction of *HAVE*, *HAD* and *HAS* with a Δ below 2

Model factors	k	df	log likelihood	AICc	Δ	w
1234567	7	13	-161.28	348.81	0	0.97
(Null)	0	3	-857.68	1721.37	1372.57	0

Note: Cutoff: $\Delta < 2$, 1 = Lemma, 2 = Following Stress, 3 = Preceding Contraction, 4 = Preceding JP, 5 = Preceding TP, 6 = Quartile, 7 = Speaker Syllable Rate

Table 35. IAS Contraction of *HAVE* Multimodel Inferencing Results

Predictor	β	σ	LoCI	HiCI	p	Cumul Prob
(Intercept)	1.092	0.054	0.054	20.358	0.000	NA
Preceding Contraction - Y	0.484	0.026	0.026	18.644	0.000	1
Preceding NP Type:						
Auxiliary (Reference Level)						1
Nominal	-0.047	0.045	0.045	1.037	0.300	
Personal Pronoun	0.493	0.022	0.022	22.490	0.000	
Non-Personal Pronoun	0.000	0.056	0.056	0.001	0.999	
Preceding Transitional Probability	0.175	0.033	0.033	5.299	0.000	1
Following Transitional Probability	-0.024	0.016	-0.055	0.008	0.140	0.83
Speaker Syllable Rate	0.000	0.000	0.000	1.235	0.217	0.75

Note: reported values are coefficients with shrinkage and adjusted standard error, $n = 948$.

Table 36. Models of IAS contraction of *HAVE* with a Δ below 2

Model factors	k	df	log likelihood	AICc	Δ	w
12345	5	10	86.39	-152.54	0	0.58
(Null)	0	3	-531.43	1068.88	1221.42	0

Note: Cutoff: $\Delta < 2$, 1 = Following TP, 2 = Preceding Contraction, 3 = Preceding NP Type, 4 = Preceding TP, 5 = Speaker Syllable Rate

4.3.2.4.3. *has- 's* and *had- 'd*

Contraction results for *HAS* and *HAD* parallel closely the results for contraction of all *HAVE* inflections combined, so no further tables of results are presented here. For *HAS*, Preceding Contraction, Preceding JP and TP as well as Utterance Quartile had high cumulative probability (over 0.60) and for *HAD*, which had fewer tokens, all of the above factors excepting Preceding TP had high cumulative probability in multimodel analyses.

4.3.2.4.4. Conclusion

We now return to the questions set forth at the beginning of this section.

Q22. Is contraction of *HAVE* influenced by the same factors that influence the word shortening of *HAVE*?

Yes and No. High probability contexts resulted in reduced variants of *HAVE*, but for word shortening, following context mattered and for contraction, preceding context mattered. For stress context, the effect was in the other direction. Preceding Stress had an effect on word shortening, but Following Stress had an effect on contraction. Both of these effects serve to promote a trochaic syllable pattern. Preceding Contraction and Syllable Rate was only relevant for contraction of *HAVE*, while Speaker Contraction Rate was only relevant for word shortening for *HAVE*. Utterance Quartile was important for both kinds of reduction. For word shortening, longer targets occurred in the final utterance quartile (more than any other quartile) and for contraction, more contraction occurred in the first quartile (more than any other quartile). This indicates that reduced variants are more likely to occur earlier in an utterance than later, but for two different

reasons. The word lengthening at the end of an utterance is due to signaling the end of an utterance or turn, or is due to less breath and intensity (cf. Lieberman, 1967). The higher contraction is likely due to the given status of the topic (additionally signaled by high rates of subject pronoun use).

Q23. Is contraction of *HAVE* influenced by the same factors that influence the contraction of *BE*?

Yes. For both *HAVE* and *BE* contraction, occurrence in the first utterance of a quartile, preceding instances of contraction, slow syllable rate, high preceding context probability and following stress syllables all lead to contraction. *BE* contraction, however, is also influenced by following context probability and *HAVE* contraction is not. There is a wider range of following TP for *BE*, than contractible *HAVE*, however, because *HAVE* can only contract in perfect constructions. Additionally, construction types (which have different contraction rates) have different average following TP rates for *BE*, which may account for the effect of following TP for *BE* words.

Q24. Are there word-specific characteristics for contraction?

No. All inflections of *HAVE* pattern similarly, even *have*, which can contract after auxiliaries (*would've*), but still contracts more often after pronouns, just as *had* and *has* do.

4.4. Discussion and Conclusion

The results in Chapter IV confirm the findings of Chapter III: function word reduction is influenced by a word's probability in context, its position in an utterance and meaning. For *BE*, contraction is influenced by construction type and for *HAVE*, it is

duration that is influenced by construction type. Table 37 provides a summary of the results.

Table 37. Buckeye Corpus Results by Dependent and Independent Variables

	Normalized Duration	BE Contraction	HAVE contraction
Probability Variables			
High Following COCA TP	↓	↓	
High Following COCA JP	↑		
High Preceding COCA TP	↑		↓
High Preceding COCA JP	↑	↓	↓
Construction			
Possessive	↑		
Modal	↑		
Perfect	↓		
Copula		↑	
Passive			
Progressive		↓	
Future		↓	
Speaker Variables			
High Contraction Rate	↓		
Fast Syllable Rate		↓	↓
Control Variables			
Preceding Stress	↓		
Following Stress		↓	↓
Following Pause			
Quartile	↑	↑	↑
Preceding Contraction	↑	↓	↓

Note: Arrows indicate a cumulative probability over 0.6 Arrows pointing up indicate the IV had a lengthening/non-reducing effect (or is associated with full forms for the contraction IV) and arrows pointing down indicate the IV had a shortening/reducing effect (or is associated with contracted forms for the contraction IV).

This section concludes Chapter IV by comparing the results from the Redford Corpus and Buckeye Corpus as well examining the effect of average probability and frequency on construction type. I argue that it is sensitivity to average high probability that results in more reduction, even in specific cases of low probability. I argue that this sensitivity means that speakers are storing phonetically detailed word representations making the production of a word (in its particular meaning) sensitive to the typical contexts in which the word tends to occur.

4.4.1. Corpus Comparison: Redford Corpus vs. Buckeye Corpus

The Buckeye Corpus is much larger than the Redford Corpus but the patterns found in function word production are remarkably similar. Because of this, we can feel confident about the results found in the previous chapter. Despite the different genres of the corpora (narratives vs. interviews) the patterns of function word reduction were remarkably similar. There were more fillers like *yknow* in Buckeye Corpus and all 40 speakers in the Buckeye Corpus had high contraction rates, but we still saw that high contraction patterns with longer word length, just on an utterance level instead of a text level. With more data, it may be that we would have seen utterance level differences in the Redford Corpus as well. In the Redford Corpus there were more effects from unexpected bigrams including words like *frog* and *turtle*. Bigrams involving story-specific words led to corpus-specific effects where bigrams had low joint probability but high transitional probability within the corpus. As the Buckeye Corpus has more data, there is naturally a larger number of distinct word types, following a power law (Zipf, 1929). Although there are certainly corpus-specific effects of specialized bigrams in the Buckeye, they are harder to spot than in the smaller Redford Corpus. Dilts (2013) found

that word informativity based on frequencies from the Buckeye Corpus better predicted content word reduction in Buckeye Corpus than word informativity based on frequencies from the larger COCA. There are probably fewer Buckeye-specific effects for function words than for content words, but they may still exist.

Because more data is available in the Buckeye Corpus, some patterns were present in this corpus that were not present in the Redford Corpus. For example, the lengthening effect of high joint probability (cf. Bell et al., 2003) was present with the larger amount of data available in the Buckeye Corpus. In the multimodel analyses presented in the Redford Corpus, many factors had high cumulative probability but still did not have a significant p value because of the shrinkage punishing the coefficients. Most of the factors with high cumulative probability in the Buckeye Corpus also had low p values. Factors reached significance due to more statistical power from the higher n . Further, we saw that the higher n and increased power resulted in more model selection certainty. There were few models with a delta value of within 2 of the best performing model and model weights were higher due to fewer probable models.

4.4.2. Construction Effects

The construction meanings for *BE* and *HAVE* with the highest frequency are the oldest (copula and possessive respectively), but not the ones associated with higher rates of reduction. Bybee (2002; 2007) posits that it is the jump in frequency of new grammatical constructions that leads to reduction of grammaticalized elements. However, in both the *BE* and *HAVE* cases, the older construction is still the most prevalent, but it is (one of) the newer constructions that has the most reduction (in terms of duration or higher proportion of contraction). If it is not an increase in frequency that leads to the

reduction, what does? The evidence in this corpus study shows that it is higher probability that results in more reduction, not higher frequency. Because a grammaticalized element occurs in more restricted conditions, the transitional probability context increases. There is a higher (and narrower) range of probability values when a word become grammaticalized and therefore restricted to certain contexts. Table 38 shows that the constructions with the most reduction have higher mean following transitional probability. The mean preceding transitional probability is more similar between the construction types, but also follows a general trend of having higher average transitional probability for constructions with more reduction.

Table 38. Mean probability contexts by Construction

Construction	Mean Post TP	Mean Pre TP
<i>BE</i>		
Copular	-1.359	-0.822
Passive	-1.052	-0.954
Progressive	-0.999	-0.860
Future	-0.668	-0.805
<i>HAVE</i>		
Possessive	-1.939	-1.326
Modal	-1.543	-1.237
Perfect	-0.855	-1.169

Note: Probability values combined from both contraction and duration data

The higher concentration of following high probability bigrams for future and progressive constructions can be seen in Figure 83 below and for perfect constructions in Figure 84 below.

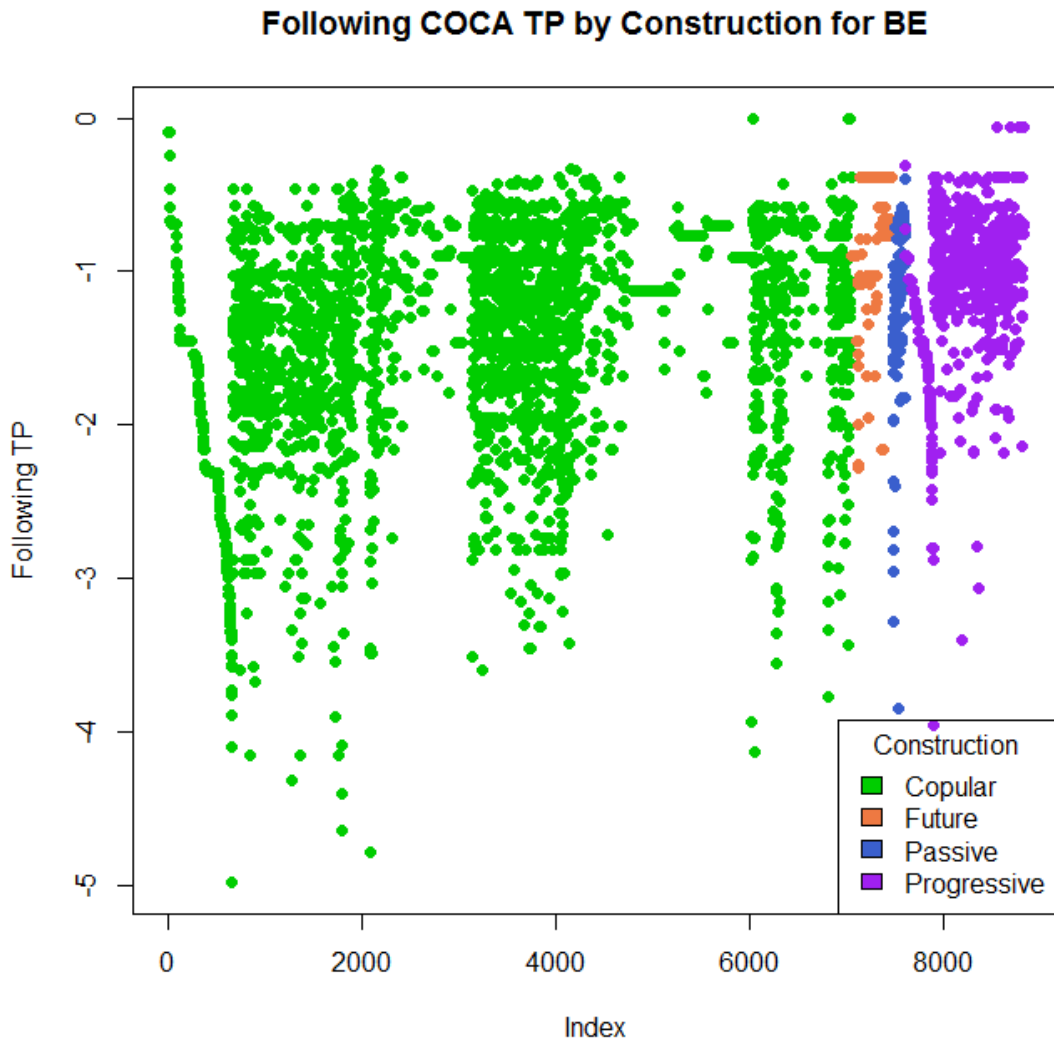


Figure 83. Following TP by Construction Type for *BE*

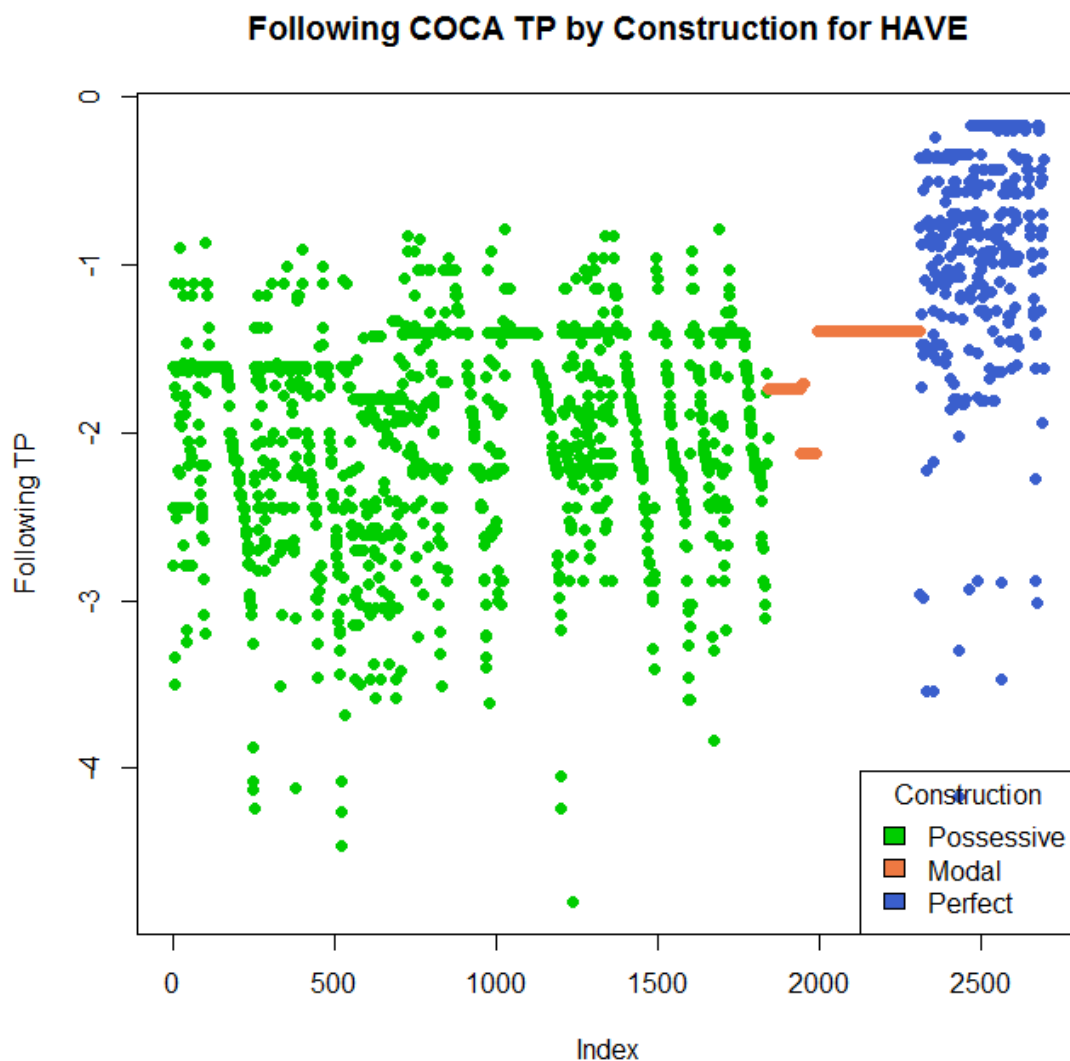


Figure 84. Following TP by Construction Type for *HAVE*

To test whether construction effects remained significant even in low probability contexts, I built models for portions of the data which had a following transitional probability below the median. For the BE dataset, to investigate construction effects for contraction, that meant instances with a following transitional probability below -1.056 (mean was -1.192). This left 4523 cases, or about half of the cases, with 3947 copula constructions, 41 future constructions, 60 passive constructions and 448 progressive

constructions. Even in these low transitional probability contexts, the main effects from the model of the full dataset remained significant. For Construction, the progressive construction continued to show significantly more contraction than the copula construction. Additionally, using the full *BE* contraction dataset, I built a logistic fixed effects regression model with following transitional probability as the sole fixed effect. I then used the residuals from this model (error leftover after taking following TP into account) as the dependent variable for a mixed effects regression model using all of the main effects from Table 27. The future and progressive constructions continued to show significantly higher contraction than the copula. All other main effects remained significant as well. For the *HAVE* dataset, to investigate construction effects for duration, I looked at the subset of cases with a following transitional probability below the median of -1.609 (mean was -1.7114). This left 1489 cases, or a bit over half of the original dataset. This dataset consisted of 1297 possessive constructions, 152 modal constructions and 40 perfect constructions. Even for this small amount of perfect constructions, the effect of construction was still significant with perfect constructions having significantly shorter *HAVE* normalized durations than other meanings of *HAVE*. Additionally, using the full *HAVE* duration dataset, I built a logistic fixed effects regression model with following transitional probability as the sole fixed effect. I then used the residuals from this model as the dependent variable for a mixed effects regression model using all of the main effects from Table 24. The targets in the perfect construction were still significantly shorter than targets in other constructions.

Seyfarth (2014) finds word informativity given the following word (average following contextual [transitional] predictability) captures word shortening effects

beyond local contextual predictability, frequency or segment count for content words in the Buckeye Corpus and the Switchboard Corpus (Godfrey and Holliman, 1997) using the NXT Switchboard Annotations (Calhoun et al., 2009). He finds that word informativity and local predictability have similar effect sizes. The current study shows that the average high probability of a construction leads to reduction or contraction even in specific low probability contexts. If speakers reduce even in particular low probability contexts, then they must be sensitive to broader patterns of use of the words in question. Bybee and Torres Cacoullos (2008) (among others including Ernestus, 2014; Goldinger, 1998; Johnson, 2007; Pierrehumbert, 2001, 2002) argue that this effect provides evidence for an exemplar account, where words that tend to occur in reduced contexts are associated with larger numbers of reduced representations than words that do not tend to occur in reduced contexts, making a reduced exemplar more likely to be selected as a target for production, even when the specific context is not one that would normally be associated with a reduced target. The exemplar account is concerned primarily with the previous experience of the speaker (also cf. Seyfarth, 2014 for a nice discussion of other production-based accounts of this finding). Some other usage-based accounts make listener-accommodation a primary consideration. Under these models, the reduction here could be accounted for by speakers knowing that listeners expect reduced variants in certain situations (short words for perfect auxiliaries, contracted forms for progressive and future auxiliaries) or unreduced variants in other situations, and then using the expected form even when the specific probability context would make that choice unlikely. Or, it may be that over time speakers use reduced forms for these contexts and learn that they are still understandable by their listeners while producing reduced variants

and that it is not necessary to go through more articulatory effort for these forms, even in low probability contexts where normally more articulatory effort would be required to be understandable (Jaeger and Ferreira, 2013; Jaeger, 2013).

In the grammaticalization literature, it is emphasized that words do not grammaticalize on their own, they grammaticalize in specific contexts (Bybee, 2002; Diewald, 2002; Heine, 1993; Hopper and Traugott, 1993; Traugott, 2003 *inter alia*). The average backward following transitional probability is a means of understanding why reduction may occur in these new contexts. When a word can be used in a new context, and its meaning is limited to that particular context, then it is part of a new construction. When the contexts a word can occur in are more limited, it becomes more predictable in those contexts. For auxiliaries in particular it is backward following transitional probability that matters. Take for example the three *HAVE* constructions. When *HAVE* is in the possessive construction, many words may follow it as part of that possessive construction including determiners, adverbs, etc. and many nouns, a large open class. This means that *HAVE* as a possessive is highly unlikely to be predictable given any of those words. When *HAVE* is used as a semi-modal auxiliary, it must occur before *to*. Although that is a very limited context, there are still many words that can occur before *to* and therefore the *backward* transitional probability is still not extremely high. When *HAVE* is used as a perfect auxiliary, it occurs before past participles, and sometimes adverbs or negator words (*he has not been to the park recently*). In the cases of adverbs or words like *not*, *HAVE* is not predictable as the word occurring before them. However, in the case of the past participles, *HAVE* is highly predictable because one of the few words that can occur, and occurs with any regularity, before past participles is *HAVE*

(and *BE*, in the case of passives). Further, the past participles that are part of passives constructions do not completely overlap with past participles that are part of perfect constructions. Past participles are sometimes used as adjectives as well, but as is captured in Figure 84 above, *HAVE* is highly likely to occur before the kinds of past participles that occur in perfect constructions.

4.4.3. *BE* – Contraction vs. *HAVE* – Duration Shortening

Many of the patterns for *BE* and *HAVE* productions are similar: high preceding joint probability resulting in contraction (where licensed), high following transitional probability resulting in shortening and contraction, following stressed syllables leading to shorter productions, among others. However, there are some particularities of each word form.

For *BE* words, both contraction and shortening are available options for reduction. It seems from the data from the Buckeye and Redford Corpus Studies, that construction differences are represented by variations in proportion of contraction, but that duration is not used to signal differences in *BE* by construction type. Even for *BE* words that cannot contract, such as *was* and *were*, we do not see any shortening by construction. But for *HAVE*, speakers can only contract perfect auxiliaries. The Buckeye Corpus shows that speakers do contract perfect auxiliaries fairly often, and do so in high probability contexts more than in low probability contexts. Despite being able to contract perfect *HAVE*, speakers still shorten perfect *HAVE* when they do produce the full variant. Why does *BE* contract but not shorten in all constructions, but *HAVE* contracts and shortens, but only in one construction? It may be that there is a floor for word shortening. On average, all the *BE* words are shorter than the *HAVE* words and the range of duration values is narrower

(cf. Table 19). It may be that there are not enough perceptible differences by construction type for shorter durations to become associated with a construction. If this were to be the case, we would still need to account for why shorter durations are still associated with higher probability contexts. A processing account of word shortening (fast access or expectation of listener comprehension even in a degraded signal) would need to exist alongside a usage-based account where certain reduction patterns become associated with constructions (contraction with some *BE* construction types, duration for perfect constructions). Another explanation may be that contraction is so strongly associated with *is*, *am* and *are* inflections of *BE* as a reduction strategy, that failing to use duration shortening has expanded to other inflections of *BE* like *was* and *were*.

In any case, this detailed study of two words has shown that there are word-specific reduction strategies in addition to general word reduction strategies, just as there are speaker-specific reduction strategies. A close examination of these words, instead of an amalgamation of all words (or all content words) in a corpus, lets us know that as researchers we need to be aware of the specific behavior of words, and that some of this behavior could not be captured by a simple random effect term of “word.” Different inflections of *BE* and *HAVE* also have special behavior, such as extremely high rates of *am* contraction. Therefore, I advocate the close examination of the effects of particular words in a corpus, in conjunction of studies of reduction collapsing many word types together.

CHAPTER V

CONCLUSION

5.1. Introduction

Several studies of probabilistic reduction have examined the impact of frequency and or probability on the production of words. Studies such as Gahl (2008) show clearly that the more frequent word of a homophone pair (*time* vs. *thyme*) is the shorter one. This shows that frequency is a property of a lemma, not just its phonological form, contra Levelt et al. (1999) and Newmeyer (2006). It is not simply the repetition of a word's phonological form over and over that leads to fluent, practiced productions (cf. Bybee, 2002; Kapatsinski, 2010). People know when a word should be more reduced (when it is predictable) and when it should not. However, Gahl (2008) examines only content words homophone pairs with different orthographic spellings. Function words and content words homophonous with function words were expressly excluded. Seyfarth (2014) showed that low average informativity of words (not just homophones) results in shorter durations, controlling for segment count, syllable count and frequency. However, he too excluded function words from his analyses. Previous studies examining function words, such as Jurafsky (2002) and Bell et al. (2003) found that once factors such as speaking rate, segmental context, pitch accent and contextual predictability are accounted for, there is no frequency effect for different meanings of words such as *to*, *that*, *of*, *you*, *I*, *and*, *the*, *a* and *it*. Krug (1998) looks at the contraction of *BE* and *HAVE*, but focuses on bigram frequency (joint probability) rather than meaning differences, meaning frequency or conditional probability on contraction.

The present study focuses on two words in particular, with different meanings that have different meaning frequencies. All meanings of *BE* are function words, allowing a controlled look at pronunciation differences among various grammatical meanings of the same word. Some meanings of *HAVE* are grammatical, but the possessive meaning is lexical, allowing an examination of grammatical vs. lexical meanings of the same word form. It does not matter if one considers these meanings to be polysemous because of the historical connection, or homonymous because of the present day differences in meaning. I show that the relative frequency of these meanings is in fact not the determining factor in production differences, contra Gahl (2008) for content words, but I also show that there are production differences according to different meanings of function words, contra Bell et al. (2003). The predictability of the particular token under investigation in its particular context is very important, suggesting an online component to reduction (Gregory et al., 1999; Jurafsky et al., 2001). However, average probability of a word's meaning across contexts is also as important, suggesting that phonetic detail is stored with forms linked to meanings a.k.a. constructions (Bybee, 2002; Pierrehumbert, 2002; Raymond and Brown, 2012; Torres Cacoullos, 1999). The detailed examination of two particular words in the present study also has shown speaker-based differences in reduction strategy patterns, word-based differences in reduction strategies, as well as a developmental trajectory of function word production in child speech. Future studies should focus on other function words to determine how much of the findings are restricted to these words, but there is no reason *a priori* that the findings would not be generalizable. Corpus studies that examine polysemous meanings of words or words that are homophonous while also being homonymous are costly in time as they require hand-

coding. However the time spent here shows that probabilistic reduction and average probability effects are not restricted to content words, showing a further similarity in content and function word production, also contra Levelt et al. (1999).

In the rest of this chapter I summarize the findings on production of *HAVE* and *BE* function words in child speech, child-directed speech and inter-adult speech. I will discuss the differences and similarities in reduction strategies between children and adults, the differences over time in child speech and the differences in adult speech directed at children vs. other adults. I will discuss what I take this to mean for theories about speech production and speaker behavior.

5.2. Child and Adult Reduction Strategies

Children and adults alike contract words in probable contexts and shorten words in probable contexts. However, adults use a consistent information compression strategy: word shortening or contraction, which is seen at the narrative level in the Redford Corpus, which interacts with speaking rate. When speaking slowly, adult speakers are more likely to contract than to shorten their words. Children develop an information compression strategy around age seven, or earlier when they have other language competence skills.

5.3. Child Reduction Strategies Over Time

Children's function word production becomes more adult like over time. Children produce shorter function words (in both raw duration and normalized duration) as they get older and they contract more often. These behaviors are consistent with deemphasizing function words, that is, deemphasizing the importance of words that are predictable to listeners. Children also have a stronger adherence to an information

compression strategy over time. They show more strongly a preference to contract (and not shorten) function words or to shorten (and not contract) function words as they get older. At the same time, children show less variability in their behavior over time. When examining the variability in the random forest analyses by grade group, the youngest grade groups often patterned against the older grade groups. Younger children have many differences in their proficiency levels, but these differences level out to a great extent over time.

The results of this study strongly support that phonetic reduction is pervasive in spontaneous speech (Cruttenden, 1994; Dalby, 1986; Dilts, 2013; Johnson, 2004; 1994; Shockey, 2003) and that this pervasive reduction starts early (already at age five), even while children are still working on getting good at reduction. That is, children are still in the process of fine-tuning their motor skills from ages five to ten in that they are still getting faster, they are still developing their ability to produce differences between strong vs. weak syllables (Allen and Hawkins, 1980; Ballard et al., 2012; Redford and Sirsa, 2011). Despite this, probabilistic reduction effects are already present in their speech. Word shortening is occurring in highly probable contexts, even when the mechanisms to shorten are still under development. The present study is the first study to investigate probabilistic reduction in child speech. Although I investigate probabilistic reduction effects for only a few words (*is, am, are, was, were, had, has, have*), this study suggests that probabilistic reduction should be happening through the lexicon at a remarkably young age. Research shows that sensitivity to transitional probability helps children bootstrap language learning and determine word boundaries (Kuhl, 2004; Saffran et al., 1996 *inter alia*) and may be a domain general learning mechanism (Ellis and Ferreira-

Junior, 2009; Kirkham et al., 2002). Therefore, it should be no surprise that children are paying attention to transitional probability at the word level and that transitional probability affects their speech production. Further study is needed to determine at what age we start to see probabilistic reduction in child speech.

5.4. Caregiver v. Inter-Adult Speech Reduction Strategies

Results from this study showed that when speaking to their young children, caregivers did not speak any slower, but that they spent proportionally more time on function words than when speaking to older children. This shows that when speaking with younger children, caregivers use a steadier rate of words, but consequently a more variable rate of information transfer, where more equal amounts of time are spent on informative and uninformative words. When speaking with older children, they use a more steady rate of information transfer, spending more time on important (lexical) information and less time on unimportant (grammatical) information, just as adults do when speaking to other adults.

A speaker-internal account of word reduction argues that speakers reduce words because they are easily accessible from the lexicon due to their high frequency and probability, and therefore can be retrieved quickly, meaning the speaker does not need to slow down while planning for the next word (cf. Ernestus, 2014). A listener-oriented account argues that because high frequency and high probability words are predictable for listeners, and speakers are aware of that, then speakers do not feel they need to spend any extra articulatory effort on uninformative, predictable words. A listener-oriented account, then, includes a model of the interlocutor for speech production (Galati and Brennan, 2010).

The results from child-directed speech in the current study suggest that speakers modulate their behavior based on characteristics of the listener. This is consistent with a listener-oriented account. However, the modulation in question does not have to be online, driven by in-the-moment modeling of the listener. Rather, the listener adaptation may be on the timescale of learning, with the speaker developing a style for speaking to a particular interlocutor or class of interlocutors (namely, five year-old children). The data from child-directed speech shows that there is more reduction of function words in caregivers' speech over time (as children get older). The function words remain equally accessible for caregiver-speakers when they are talking to five year-olds and when they are talking to ten year-olds. Therefore the longer productions associated with speaking to five year olds must be due to accommodation, whether online or longer-term.

5.5. Usage-Based Speech Production

To accommodate the results found in the present study, a theory of speech production must be able to account for the following four facts:

- A) The degree of reduction differs for the same string of phonemes, beyond reduction due to speech rate
- B) The degree of reduction for a word differs based on the specific transitional probability of that word and the word following it
- C) The degree of reduction for a wordform-meaning pairing (construction) differs based on the average transitional probability of that construction given the words following it

D) The degree of reduction for a word as described in C is not due to differences in raw frequency of that word

Abstractionist models of speech production assume that word (and morpheme) representations in the mental lexicon consists of strings of abstract symbols, often phonemes (Chomsky and Halle, 1968; Levelt et al., 1999; Prince and Smolensky, 2004). Although specific instances of production differ, they come from one representation, passed to a phonetic implementation system where rules or constraints, even optional rules or variable constraints (cf. Boersma, 1998; Cedergren and Sankoff, 1974), result in the phonetic detail of the specific instance of production. A key consequence of these models of speech production is that the meaning of the word should have no bearing on the pronunciation of any particular string of phonemes and that reduction associated with one word should be inherited by any homonyms (Gahl, 2008; Levelt et al., 1999). Abstractionist models of this sort fail to account for finding A, as shown in many studies cited in Chapter II, as well.

Pierrehumber (2002) and Ernestus (2014) explain how finding B could still be accommodated in abstractionist models of speech production. Although Pierrehumbert (2002) argues for an exemplar model with long-term memories of linguistic events (including speaker and phonological information) which affect phonological encoding and the representation of a lexical network, she argues that reduction due to high frequency and high probability effects could be theoretically implemented as due to on-line modifications of speech. In a modular feedforward model, “when a lexeme is retrieved and loaded into the phonological buffer, assume that a gradient value representing the ease of retrieval is passed to the buffer as a quantitative attribute of the

Prosodic Word node. This parameter would control, or rather play a part in controlling, an overall parameter of articulatory clarity and effort” (Pierrehumbert, 2002:104).

Ernestus (2014:31), who also argues for a model of speech processing which has storage of different pronunciation variants as well as procedural phonetic implementation, explains how this could work (although it is not clear why it is predictability of the current word rather than that of the upcoming word that matters on this account):

Interestingly, reduction degree appears to be correlated especially with the predictability of the word given the following word rather than the preceding word. These predictability effects can be well accounted for with the assumption that more predictable words are easier to plan and to retrieve from the mental lexicon (e.g., Jescheniak and Levelt, 1994) and therefore do not require speakers to slow down: While planning highly predictable words, speakers can continue speaking as fast as they would like to. The higher speech rate with which highly predictable units can be produced would be responsible for their higher reduction degree (see e.g., Pluymaekers et al., 2005a; Bell et al., 2009).

Abstractionist models that allow for the implementation of variable production due to high context probability still fail to accommodate finding C. Perfect auxiliaries, even when they have low backward following transitional probability, are still short.

Progressive and future auxiliaries, even when they have low backward following transitional probability, are still likely to be contracted. Ernestus (2014) also reviews some studies that show that some aspects of variation in word reduction cannot be explained by online processes. Words with the same phonological, prosodic and predictability characteristics should not differ in their degree of reduction if done through online phonetic processes, but they do. The bigram *I think*, although consistent in its phonological makeup and bigram probability, differs in realization when it conveys a pragmatic meaning rather than a lexical one (Local, 2003; see also Bybee and Scheibman, 1999). The realization of the Dutch word *eigenlijk* ‘actually’ differs

depending on whether the word signals a contrast with a previous proposition or a speaker's assumption of the listener's expectations (Plug, 2005).

Because of these findings, usage-based models of speech production are more convincing in accounting for the variable productions. However, the reduction associated with different word forms cannot be due only to a word's raw frequency (as seen in Aylett and Turk, 2006; Bybee, 2001; Gahl, 2008) due to finding D. The words with the highest degree of reduction in this study are not the most frequent, in fact they are far less frequent than their source constructions.

A usage-based theory of speech production is one in which generalizations, representations and categories emerge through language use. There are no strong boundaries between levels of representation (i. e., grammar v. the lexicon, the lexicon v. phonological system), rather all levels of representation are interconnected in a network (Bates and Goodman, 1997, Bybee, 2001; Langacker, 1987; 2000). Factors of usage, such as word frequencies, word co-occurrence and stochastic variability influence language representation and these representations can change over time due to changes in experience. Below I discuss two usage-based models and how the findings from the present study can be understood (or not) under these frameworks. I discuss an exemplar model and a schema-based development model and how they relate to acquisition and the present data.

There are many examples of exemplar-based models in speech production literature (Bybee and Torres Cacoullos, 2008; Goldinger, 1998; Johnson, 2007; Munson, 2010; Pierrehumbert 2001; 2002 *inter alia*). In exemplar-based models, speakers store memories of word productions. All pronunciation variants are found together in a word

cloud in the mental lexicon. When a speaker plans word production, they select an episodic memory as the basis for their production. Words that are more frequently reduced have more reduced representations in the word cloud, so that when a speaker goes to produce a word, in turn they are more likely to select a reduced representation as a basis for production and then produce a reduced variant. Episodic memory of words includes semantic and syntactic properties (and perhaps also social contexts), so speakers will activate a memory (or cloud of memories) most closely relevant to the current situation for speech planning. As speakers build their cloud of word memories, they become aware of the patterns in variability through statistical learning (Pierrehumbert, 2003). From this sensitivity to the patterns in word clouds emerges speaker awareness of phonological categories (Pierrehumbert and Gross, 2003) and could also emerge awareness of constructional categories, such as future and progressive constructions tend to have contracted auxiliaries, etc.

Foulkes and Docherty (2006) argue for an exemplar model that integrates issues of language acquisition and sociophonetic variability. They argue that a lexical representation should include detailed acoustic traces of every experience of a word that an individual has had, along with information about who was speaking and what the speaker's voice sounded like so that no human utterance occurs absent of indexing social factors. Foulkes and Docherty (2006) also assume that there will be abstractions over exemplars and associations with different social indexical information (presumably along with structural and neighboring context information) and that speakers will become aware of (or at least sensitive to) acoustic proprieties of the speech that highlight social indexation of different kinds. In early acquisition, they posit, it is likely that children will

have three basic groups that emerge through their experience with their families: adult males, adult females and children, based on nature-based divisions like F0 and formant frequencies. Although the child may initially form groups based on specific people (father, mother, sibling), they later will be able to make generalizations to kinds of people from their experience with specific people. Some kinds of variation associated with groups of people would be easier to learn, variation that has a biological basis for instance, but through enough experience and exposure, children would learn arbitrary associations between groups of people and sociophonetic patterns. As children get older, and as their social world becomes wider and more complex, they may pay attention to speech they see as better representing themselves (a boy paying attention to adult male speech despite its lower frequency of exposure) or consider certain types of speech to be more important because it is closer to what they are able to produce (school aged children beginning to sound more like their friends than their parents, cf. Kerswill and Williams (2000)). While children are learning the social associations of speech, they are also learning the lexical and structural associations that result in reduction in this study. The associations are of a different kind, but if categories of complex social variability can emerge through exposure to speech, then other categories of complex variability can emerge as well and indeed seem to, based on the results of this study.

The results from the present study are compatible with an exemplar based model of language production, in that there is patterned variation in pronunciation, in child as well as adult speech. In the present study we saw that the most frequent examples do not lead to more reduction, rather it is examples with high conditional probability (and average high conditional probability) that lead to reduction. Bybee and Torres Cacoullos

(2008) argue that words which often occur in reduced contexts have stronger reduced representations than words that do not tend to occur in reduced contexts. This makes a reduced exemplar more likely to be selected as a target for production, even when the specific context is not one that would normally be associated with a reduced target. Take as example, *HAVE* production. Listeners have many exemplars of possessive constructions but fewer exemplars of perfect constructions, but all of the latter tend to be short. When new perfect constructions are produced, exemplars are selected from the storage of perfect exemplars, which tend to be reduced, leading to short productions. When a speaker wants to produce a perfect construction using a series of words they have never heard before, despite its low frequency, they will still select a reduced exemplar as a basis for their production, because most perfect constructions are represented with reduced exemplars in the exemplar-cloud lexical representation. In this way, it would not be necessary for a speaker to have built up a generalization that perfect auxiliaries tend to be short; because of their exposure to short tokens, new tokens will also be short. In an exemplar model that includes abstractions over categories, a speaker would have an averaged perfect auxiliary from all of their exposures (or perhaps several averaged representations that represent different styles, speaker categories, etc. or whatever emerged as relevant). This averaged representation would be biased towards being short from repeated exposure to short examples and so when a speaker went to produce a perfect auxiliary using a string of words they have never before encountered, that production would still be short, despite the low frequency of the bigrams. Because productions differ due to specific probability of bigrams, not just averages, online reduction mechanisms would still be needed.

5.6. Grammaticalization

It is well known that one of the consequences of grammaticalization can be phonetic erosion, or reduction (Gabelentz, 1891; Lehmann, 1995; Bybee and Pagliuca, 1985; Givón, 1985; Heine, 1993; 2003; Hopper and Traugott, 1993; Bybee, Perkins and Pagliuca, 1994). It is argued that erosion can take place because of a speaker's desire to differentiate the grammaticalized words (or words in context, namely constructions) from the source words, or because the grammaticalized word increases in frequency in comparison to the source word (Bybee, 2001; 2007; Traugott, 2011). The results of the present study show that an increase in frequency cannot be the reason for the higher levels of reduction of the auxiliaries in the perfect, progressive and future constructions. However, the increase in predictability due to the narrower context in which the grammaticalized use of a word comes to occur as a result of grammaticalization (Bybee, 2002; Diewald, 2002; Heine, 1993; Hopper and Traugott, 1993; Traugott, 2003) is a good reason for the higher level of reduction in some of the more grammaticalized uses of the examined auxiliaries.

For *HAVE*, there are two constructions that have grammaticalized from the source possessive construction: perfect and semi-modal. The average following transitional probability for the two constructions is similar (cf. Table 38) but the average following transitional probability for the perfect construction is much higher. As seen in the results above, child and adult speakers alike do not shorten the semi-modal auxiliary any more than the possessive verb. However, as expected if reduction is driven by average predictability of the target word within the construction, the perfect auxiliary is significantly more likely to reduce than either the possessive or the semi-modal.

For *BE*, there are two constructions that have grammaticalized from the source copula construction: passive and progressive. The future construction has further grammaticalized from the progressive construction. The progressive and future constructions have higher levels of contraction rates in child, child-directed and inter-adult speech. These constructions also have higher average following transitional probability than the copula construction. The passive construction has higher average following transitional probability than the copula construction, but not much higher. Consequently, the passive construction is associated with more contraction in child-directed and inter-adult speech, but not significantly so. For children, the passive construction is associated with significantly less contraction than the other construction types. Children do not use the passive construction often and the passive construction, for them, may feel more formal or awkward. This shows that it is not solely something like informativity or average transitional probability that influences reduction. Other aspects associated with the construction, such as context of use (Raymond and Brown, 2012; Torres Cacoullos, 1999) and usage frequency (Alba, 2008; Bybee, 2002; 2007; Hollman & Siewierska 2007; Torres Cacoullos and Walker, 2011) also matter. However, the measure of average following transitional probability quantitatively captures the intuition that grammaticalization researchers have had for over a century that more grammatical (and therefore less informative and more predictable, a.k.a. ‘bleached’) information is more subject to reduction (or phonetic erosion) than source lexical items. This account is advantageous as it also corresponds with psycholinguistic research showing the same tendencies are found throughout language production as speakers spend less time and effort on producing items that are predictable for the listener (Arnold, 2008; Aylett &

Turk, 2004; 2006; Bell et al., 2003; Bell et al., 2009; ; Byrd, 1994; Bybee, 2001; Clark et al., 1991; Clark et al., 1987; Ernestus, 2014; Ferguson, 1977; Ferreira, 2008; Flemming, 2010; Fowler & Housum, 1987; Fox Tree and Clark, 1997; Freed, 1978; Gahl, 2008; Gahl and Garnsey, 2004; Gahl, et. al, 2012; Galati and Brennan, 2010; Gregory et al., 1999; Jaeger and Tily, 2010; Jurafsky et al., 2001; Jurafsky, 2003; Levy and Jaeger, 2007; Lindblom, 1990; Lockridge and Brennan, 2002; Munson, 2007; Pluymaekers, et al., 2005; Seyfarth, 2014; Swanson et al., 1992; van Bergem, 1995; van Son and Pols, 2003). One novel contribution of the present work is that predictability does not only influence reduction in the moment of speaking. Predictability effects also accumulate over time, so that word uses that occur in predictable contexts may also be reduced in contexts where they are less predictable due to accumulation of reduced pronunciations of that lemma in long term memory. Grammaticalization causes the grammaticalized lemma to consistently occur in specific contexts where it is relatively predictable, thus resulting in both increased online reduction of the lemma's form and increased association of the lemma with reduced form exemplars.

REFERENCES CITED

- Alba, Matt (2008). 'Ratio frequency: Insights into usage effects on phonological structure from hiatus resolution in New Mexican Spanish', *Studies in Hispanic and Lusophone Linguistics*, 1, 247-86.
- Allen, G., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology: Volume 1. Production*. (pp. 227-256). New York: Academic Press.
- Ansaldo, U., & Lim, L. (2004). Phonetic absence as syntactic prominence. Grammaticalization in isolating tonal languages. In O. Fisher, M. Norde & H. Perridon (Eds.), *Up and down the cline - the nature of grammaticalization* (pp. 345-362). Philadelphia, Amsterdam: John Benjamins.
- Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, 23(4), 495-527.
- Aylett, M., & Turk A. (2004). The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence and duration in spontaneous speech. *Language and Speech*, 47, 31-56.
- Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, 119, 3048-3058.
- Baddeley, A. (1986). Working memory. Oxford: Clarendon
- Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J., & Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics*, 39(1), 1-17.
- Baker, R. E., & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*, 52(4), 391-413.
- Ballard, K., Djaja, D., Arciuli, J., James, D., & van Doorn, J. (2012). Developmental trajectory for production of prosody: lexical stress contrastivity in children ages 3 to 7 years and in adults. *Journal of Speech, Language, and Hearing Research*, 55, 1822-1835.
- Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24(4), 527-554.
- Bard, E. G., & Aylett, M. P. (2004). Referential form, word duration, and modeling the listener in spoken dialogue. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use* (pp. 173-191). Cambridge, MA: MIT Press.

- Barry, C., Morrison, C., & Ellis, A. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency and name agreement. *Quarterly Journal of Experimental Psychology*, 50A(3), 560-585.
- Barth, D. and V. Kapatsinski. (In Press). A multimodel inference approach to categorical variant choice: construction, priming and frequency effects on the choice between full and contracted forms of *am*, *are* and *is*. *Corpus Linguistics and Linguistic Theory*.
- Bates, E., Chen, S., Tzeng, O., Li, P., & Opie, M. (1991). The noun-verb problem in Chinese aphasia. *Brain and Language*, 41(2), 203-233.
- Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher and B. MacWhinney (Eds.), *The handbook of child language* (pp. 96-51). Oxford: Blackwell.
- Bates, E., Devescovi, A., & Wulfeck, B. (2001). Psycholinguistics: A cross-language perspective. *Annual Review of Psychology*, 52(1), 369-396.
- Bates, E., Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*, 12(5-6), 507-584.
- Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, 22, 201-211.
- Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M. L., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, 113(2), 1001-1024.
- Bell, A., Brenier, J., Gregory, M. L., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92-111.
- Bellinger, D. (1980). Consistency in the pattern of change in mothers' speech: some discriminant analyses. *Journal of Child Language*, 7, 469-487.
- Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, 33(1), 73-79.
- Bock, H. (1931). Studien zum präpositionalen Infinitiv und Akkusativ mit dem *TO*-Infinitiv. *Anglia*, 55, 114-249.
- Boersma, P. (2008). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

- Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4, retrieved 4 October 2014 from <http://www.praat.org/>
- Bolinger, D. (1972). Accent is predictable (If you're a mind-reader), *Language*, 48, 633-644.
- Bolinger, D. (1975). *Aspects of Language*. New York: Harcourt Brace Jovanovich.
- Bolinger, D. (1985). Two views of accent. *Journal of Linguistics*, 21(1), 79-123.
- Bookheimer, S. Y., Zeffiro, T. A., Gaillard, W., & Theodore, W. (1993). Regional cerebral blood flow changes during the comprehension of syntactically varying sentences. *Social Neuroscience Abstract*, 19, 843.
- Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role and gender. *Language and Speech*, 44(2), 123-147.
- Braine, M. D. (1963). The ontogeny of English phrase structure: The first phase. *Language*, 1-13.
- Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, 44, 274-296.
- Bresnan, J., & J. Spencer. (2013) Frequency and variation in English subject-verb contraction. Stanford, CA: Stanford University Department of Linguistics and Center for the Study of Language and Information ms. (submitted for publication)
- Brinton, L. J. (1988). *The development of English aspectual systems*. Cambridge: Cambridge University Press.
- Brown, R. (1973). *A first language*. Cambridge: Harvard University Press.
- Brown, R., & Bellugi-Klima, U. (1964). Three process in the child's acquisition of syntax. *Harvard Educational Review*, 34, 307-18.
- Brysbaert, M., New, B., & Kueleers, E. (2012). Adding part of speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, 44(4), 991-997.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical Information-Theoretic approach*, 2nd edition. New York: Springer.
- Burrowville, J. A., & Turville-Petre, T. (1996). *A book of Middle English*. Oxford: Blackwell Publishers Ltd.

- Bush, N. (2001). Frequency effects and word-boundary palatalization in English. In J. Bybee & P. Hooper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 255-280). Amsterdam: Berlin.
- Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. Philadelphia, PA: Benjamins.
- Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261-290.
- Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford ; New York: Oxford University Press.
- Bybee, J. L., & Pagliuca, W. (1985). Cross-linguistic comparison and the development of grammatical meaning. In J. Fisiak (Ed.), *Historical semantics, historical word formation* (pp. 59-83). Berlin: Mouton de Gruyter..
- Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press.
- Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics*, 37, 575-596.
- Bybee, J., & Torres Cacoullos, R. (2008). Phonological and grammatical variation in exemplars models. *Studies in Hispanic and Lusophone Linguistics*, 1(2), 399-413.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15(1), 39-54.
- Caballero, G., & Kapatsinski, V. (2014). Perceptual functionality of morphological redundancy in Choguita Rarámuri (Tarahumara). *Language, Cognition and Neuroscience*, (ahead-of-print), 1-10.
- Calhoun, S., Carletta, J., Jurafsky, D., Niossim, M., Ostendorf, M., & Zaenen, A. (2009). NLT switchboard annotations. Technical report, Linguistic Data Consortium, Philadelphia.
- Cameron, R., & Flores-Ferrán, N. (2004). Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context*, 1, 41-65.
- Cedergren, H. J., & Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, 50, 333-355.

- Chen, S., & Bates, E. (1998). The dissociation between nouns and verbs in Broca's and Wernicke's aphasia: Findings from Chinese. *Aphasiology*, *12*(1), 5-36.
- Cheng, H. Y., Murdoch, B. E., Goozee, J. V., & Scott, D. (2007). Electropalatographic assessment of tongue-to-palate contact patterns and variability in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research*, *50*(2), 375-392.
- Chevrot, J., Beaud, L., & Varga, R. (2000). Developmental data on French sociolinguistic variable: post-consonantal word-final /R/. *Language Variation and Change* *12*: 295-319.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.
- Clark, H. H., Brennan, S. E., Resnick, L. B., Levine, J. M., & Teasley, S. D. (1991). *Grounding in communication perspectives on socially shared cognition* (pp. 127-149). Washington, DC: American Psychological Association.
- Clark, J. E., Lubker, J. F., & Hunnicutt, S. (1987). Some preliminary evidence for phonetic adjustment strategies in communication difficulty. In R. Steele & T Threadgold (Eds.), *Language topics: Essay in honour of Michael Halliday* (pp. 161-180). Amsterdam: Benjamins.
- Coates, J. (1988). The acquisition of the meanings of modality in children aged eight and twelve. *Journal of Child Language*, *15*, 425-434.
- Cortese, M., & M. Khanna. (2008). Age of acquisition ratings for 3,000 monosyllabic words. *Behavior Research Methods*, *40*(3), 791-794.
- Cruttenden, A. (1994). *Gimson's pronunciation of English* (5th ed.). London: Edward Arnold.
- Dalby, J. (1986). *Phonetic structure of fast speech in American English*. Bloomington, IN: Indiana University Linguistics Club.
- Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D. & Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature*, *380*, 499-505.
- Davis, B. L., MacNeilage, P. F., Matyear, C. L., & Powell, J. K. (2000). Prosodic correlates of stress in babbling: an acoustical study. *Child Development*, *71*, 1258-1270.
- del Prado Martín, F. M., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, *94*(1), 1-18.

- Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge: Cambridge University Press.
- Diewald, G. (2002). A model for relevant types of contexts in grammaticalization. *Typological Studies in Language*, 49, 103-120.
- Dilts, P. (2013). *Modelling phonetic reduction in a corpus of spoken English using random forests and mixed effects regression*. Edmonton, Alberta: PhD. dissertation. University of Alberta.
- Echols, C. H., & Newport, E. (1992). The role of stress and position in determining first words. *Language Acquisition*, 2, 189-220.
- Ellis, N. C., & Ferreira-Junior, F. (2009). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, 93(3), 370-385.
- Elman, J. L. (2001). Connectionism and language acquisition. In M. Tomasello & E. Bates (Eds.), *Language development: The essential readings* (295-306). Oxford: Basil Blackwell.
- Embick, D., Marantz, A., Miyashita, Y., O'Neil, W., & Sakai, K. L. (2000). A syntactic specialization for Broca's area. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 6150-6154.
- Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, 142, 27-41.
- Ernestus, M., Baayen, R. H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81, 162-173.
- Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39(3), 253-260.
- Fenk, A., & Fenk, G. (1980). Konstanz im Kurzzeitgedächtnis-Konstanz im sprachlichen Informationsfluß. *Zeitschrift für experimentelle und angewandte Psychologie*, 27, 400-414.
- Ferguson, C. A. (1977). Baby talk as a simplified register. In C. E. Snow and C. A. Ferguson (Eds.), *Talking to children* (209-235). Cambridge: Cambridge University Press.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16, 477-501.

- Ferreira, V. S. (2008). Ambiguity, accessibility, and a division of labor for communicative success. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 210-246). Amsterdam: Elsevier.
- Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40, 296-340.
- Finegan, E. (1995). Subjectivity and subjectification: An introduction. In D. Stein and S. Wright (Eds.), *Subjectivity and subjectification: Linguistic perspectives* (pp. 1-15). Cambridge: Cambridge University Press.
- Fischer, O. (1994). The development of quasi-auxiliaries in English and changes in word order. *Neophilologus*, 78, 137-164.
- Flemming, E. (2010). Modeling listeners: Comments on Pluymaekers et al. and Scarborough. In C. Fougeron, B. Kühnert, M. D'Imperio and N. Vallée (Eds.), *Laboratory Phonology 10* (pp. 587-606). Berlin: Mouton.
- Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of experimental psychology: Learning, Memory, and Cognition*, 10(4), 680-698.
- Foulkes, P., & Docherty, G. (2005). The social life of phonetics and phonology. *Journal of Phonetics*. 34. 409-438.
- Foulkes, P., Docherty, G., & Watt, D. (2005). Phonological variation in child-directed speech. *Language*, 81(1), 177-206.
- Fosler-Lussier, E., & Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication*, 29, 137-158.
- Fowler, C. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, 31(4). 307-319.
- Fowler, C., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26. 489-504.
- Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition* 62:151-167.
- Frank, A., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. *Cogsci. Washington, DC: CogSci*.
- Freed, B. F. (1978). *Foreign talk: A study of speech adjustments made by native speakers of English in conversation with non-native speakers*. Philadelphia: PhD. dissertation. University of Pennsylvania.

- Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph Drawing by Force-Directed Placement. *Software: Practice and Experience*, 21(11), 1129-1164.
- Gabelentz, G. (1891). *Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Leipzig: T. O. Weigel.
- Galati, A., & Brennan, S. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language* 62(1). 35-51.
- Gahl, S. (2008). "Time" and "thyme" are not homophones: Word durations in spontaneous speech. *Language*, 84, 474-496.
- Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, 80(4), 748-775.
- Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66, 789-806.
- Garnica, O. (1977). Some prosodic and paralinguistic features of speech to young children. In C. E. Snow and C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 63-88). Cambridge: Cambridge University Press.
- Gathercole, S. E., & Baddeley, A. (1993). *Working memory and language*. New York: Taylor & Francis.
- Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, 63, 1816-1827.
- Gay, T. (1981). Mechanisms in the control of speaking rate. *Phonetica*, 27, 44-56.
- Gallaway, C., & Richards, B. J. (Eds.). (1994). *Input and interaction in language acquisition*. Cambridge: Cambridge University Press.
- Garside, R. (1996). The robust tagging of unrestricted text: the BNC experience. In J. Thomas and M. Short (Eds.), *Using corpora for language research: Studies in the honor of Geoffrey Leech* (pp. 167-180). London: Longman.
- Gee, J. (1985). An interpretive approach to the study of modality: What child language can tell the linguist. *Studies in Language*, 9, 127-229.
- Gilhooly, K., & Logie, R. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, 12(4), 395-427.

- Givón, T. (1985). Iconicity, isomorphism and non-arbitrary coding in syntax. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 187-219). Amsterdam: John Benjamins.
- Givón, T. (1991). Serial verbs and the mental reality of 'event': Grammatical vs. cognitive packaging. *Approaches to grammaticalization, 1*, 81-127.
- Godfrey, J. J., & Holliman, E. C. (1997). Swichboard-1 release-2. Technical report, Linguistic Data Consortium, Philadelphia.
- Goffman, L. (1999). Prosodic influences on speech production in children with specific language impairment and speech deficits. *Journal of Speech, Language, and Hearing Research, 42*, 1499-1517.
- Goffman, L. (2004). Kinematic differentiation of prosodic categories in normal and disordered language development. *Journal of Speech, Language, and Hearing Research, 47*, 1088-1102.
- Goffman, L., & Malin, C. (1999). Metrical effects on speech movements in children and adults. *Journal of Speech, Language, and Hearing Research, 42*, 1003-1015.
- Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure Constructions*. Chicago, IL: University of Chicago Press.
- Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 195*(2), 251-279.
- Goldman-Eisler, F. (1957). Speech production and language statistics. *Nature, 180*, 1487.
- Goldsmith, J. (2007). Probability for linguists. *Mathématiques et sciences humaines. Mathematics and social sciences, 180*, 73-98.
- Gordon, P., & Chafetz, J. (1991). Verb-based vs. class-based accounts of actionality effects in children's comprehension of the passive. *Cognition, 36*, 227-254.
- Gregory, M. L., Raymond, W., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society, 35*. 151-166.
- Gregory, M. L., Healy, A. F., & Jurafsky, D. (2001). *Common ground in production: Effects of mutual knowledge on word duration*. Unpublished manuscript, University of Colorado at Boulder.
- Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics 19*(1). 1-33.

- Heine, B. (1993). *Auxiliaries*. Oxford: Oxford University Press.
- Heine, Bernd (2003) 'Grammaticalization.' In Brian D. Joseph & Richard D. Janda (Eds.) *The Handbook of Historical Linguistics* (pp. 575-601). Oxford: Blackwell.
- Hochberg, J. (1986). Functional compensation for /-s/ deletion in Puerto Rican Spanish. *Language*, 62, 609-621.
- Hopper, P., & Traugott, E. C. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.
- Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. (2006). Survival ensembles. *Biostatistics*, 7(3), 355-373.
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651-674.
- Hundt, M. (2004). The passival and the progressive passive: A case study of layering in the English aspect and voice systems. In C. Lindquist, C. and H. Mair. (Eds.), *Corpus approaches to grammaticalization in English* (pp. 79-120). Amsterdam: Benjamins.
- Israel, M., Johnson, C., & Brooks, P. J. (2000). From states to events: the acquisition of English passive participles. *Cognitive Linguistics*, 11, 103-129.
- Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, 58(2), 541-573.
- Izura, C., & A. Ellis. (2002). Age of acquisition effects in word recognition and production in first and second languages. *Psicológica*, 23, 245-281.
- Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language variation and change*, 21(2), 233-256.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61, 23-61.
- Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology*, 4, 230.
- Jaeger, T. T., & Ferreira, V. (2013). Seeking predictions from a predictive framework. *Behavioral and Brain Sciences*, 36(4), 31-32.
- Jaeger, T. F., & Tily, H. (2010). On language 'utility': Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(3), 323-335.

- Jescheniak, J.D., & Levelt, W.J.M. (1994). Word frequency effects in speech production: retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 824-843.
- Johnson, K. (1994). Massive reduction in conversational American English. In *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium* (pp. 29-54). Tokyo, Japan: The National International Institute for Japanese Language.
- Johnson, K. (2007). Decisions and mechanisms in exemplar-based phonology. In Sole, M., Beddor, P., & Ohala, M. (Eds.), *Experimental Approaches to Phonology. In Honor of John Ohala* (pp. 25-40). Oxford: Oxford University Press.
- Jones, C. (1972). *An Introduction to Middle English*. New York: Holt, Rinehart and Winston, Inc.
- Jurafsky, D., Bell, A., Gregory, M. L., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee, and P. Hopper. (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam: Benjamins.
- Jurafsky, D., Bell, A., Gregory, M. L., & Raymond, W. D. (2002). The role of the lemma in form variation. In Gussenhoven, C., & Warner, N. (Eds.), *Laboratory Phonology* (pp. 1-34). Berlin: Mouton de Gruyter.
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159-207.
- Kapatsinski, V. (2005). Measuring the relationship of structure to use: Determinants of the extent of recycle in repetition repair. *Berkeley Linguistics Society*, 30, 481-92.
- Kapatsinski, V. (2010). Frequency of use leads to automaticity of production: Evidence from repair in conversation. *Language & Speech*, 53(1), 71-105.
- Kehoe, M. (2001). Segmental timing of young children and adults. *International Journal of Speech-Language Pathology*, 12, 284-294.
- Kehoe, M., Stoel-Gammon, C., & Buder, E. H. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research*, 38, 338-350.
- Kempe, V., Schaeffler, S., & Thoresen, J. (2010). Prosodic disambiguation in child-directed speech. *Journal of Memory and Language*, 62(2), 204-225.
- Kendall, T. (2009a.) parse_textgrid_transcript.praat, retrieved 26 May 2011 from http://ncslaap.lib.ncsu.edu/tools/scripts/parse_textgrid_transcript.praat

- Kendall, T. (2009b). vowel_capture_aug09.praat, retrieved 26 May 2011 from http://ncslaap.lib.ncsu.edu/tools/scripts/vowel_capture_aug09.praat.
- Kendall, T. (2011). english_syllable_counter.R, retrieved 24 May 2011 from web/personal communication.
- Kent, R. D., & Forner, L. L. (1980). Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics*, 8, 157-168.
- Kerswill, P. E., & Williams, A. (2000). Creating a new town koine: Children and language change in Milton Keynes. *Language in Society*, 29, 65-115.
- King, J. W., & Kutas, M. (1998). Neural plasticity in the dynamics of human visual word recognition. *Neuroscience Letters*, 244(2), 61-64.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83(2), B35-B42.
- Kovac, C., & Adamson, H. D. (1981). Variation theory and first language acquisition. In D. Sankoff and H. Cedergren (Eds.), *Variation Omnibus* (pp. 403-410). Edmonton, Alberta: Linguistic Research, Inc.
- Kowal, S., O'Connell, D. C., & Sabin, E. J. (1975). Development of temporal patterning and vocal hesitations in spontaneous narratives. *Journal of Psycholinguistic Research*, 4, 195-207.
- Krug, M. (1998). String frequency: a cognitive motivating factor in coalescence, language processing, and linguistic change. *Journal of English linguistics*, 26, 286-320.
- Krug, M. (2000). *Emerging English modals, a corpus based study of grammaticalization*. Berlin: Mouton de Gruyter.
- Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in cognitive sciences*, 14(7), 293-300.
- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5, 831-843.
- Kuperberg, G. E., McGuire, P. K., Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Wright, I. C., Lythgoe, D. J., Williams, S. C. R., & David, A. S. (2000). Common and distinct neural substrates for pragmatic, semantic and syntactic processing of spoken sentences: an fMRI study. *Journal of Cognitive Neuroscience*, 12, 120-133.

- Kuperman, V., & Bresnan, J. (2012). The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language*, 66, 588-611.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods*, 44(4), 978-990.
- Kurumada, C., & Jaeger, T. F. (2013). Communicatively efficient language production and casemarker omission in Japanese. In *CogSci 12: The 34th Annual Meeting of the Cognitive Science Society. Sapporo, August 2012*.
- Kuznetsova, A. (2014). lmerTest: Tests in Linear Mixed Effects Models. Version 2.0-20. Package for R. <http://cran.r-project.org/web/packages/lmerTest/index.html>
- Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, 45, 715-62.
- Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford university press.
- Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow and S. Kemmer (Eds.), *Usage-based models of language* (pp. 1-63). Stanford, CA: CSLI Publications.
- Langacker, R. W. (2011). Grammaticalization and Cognitive Grammar. In B. Heine & H. Narrog, (Eds.), *The Oxford handbook of grammaticalization* (pp. 79-91). Oxford: Oxford University Press.
- Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, 105, 1455-1468.
- Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in Contemporary English*. Cambridge: Cambridge University Press.
- Lehmann, C. (1995). *Thoughts on grammaticalization*. München: Lincom Europa.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-75.
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In *Paper presented at the twentieth annual conference on neural information processing systems*.
- Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6(3), 172-187.

- Lieberman, P. (1967). *Intonation, perception, and language*. Cambridge, MA: MIT Press.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle and A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403-439). Dordrecht: Kluwer.
- Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, 9, 550-557.
- Lord, C., Berdan, R., & Fender, M. (2009). Distinguishing function words from content words in children's oral reading. In D. Lardiere, R. P. Leow and H. Campos (Eds.), *Little words: Their history, phonology, syntax, semantics, pragmatics, and acquisition* (pp. 37-46). Washington, D. C.: Georgetown University Press.
- MacKenzie, L. (2012). *Locating Variation Above the Phonology*. Philadelphia: PhD dissertation. University of Pennsylvania.
- Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15(1), 19-44.
- Major, D. (1974). *The acquisition of modal auxiliaries in the language of children*. The Hague: Mouton.
- Malone, D., & Sullivan, W. (2005). Guesswork is not a substitute for entropy. *Proceedings of the Information Technology & Telecommunications Conference*.
- Manning, C., & Schütze, H. (1999). *Collocations. Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.
- Massey, J. (1994). Guessing and entropy. *Proceedings of IEEE International Symposium on Information Theory*.
- Mathieu-Reeves, D., & Redford, M. A. (2009). *Is is always is? Phonetic effects of grammaticalization*. Paper presented at *Current Trends in Grammaticalization Research*. Groningen, Netherlands, October 8-9.
- Mayer, M. (1967). *A Boy, a Dog and a Frog*. New York: Dial Press.
- Mayer, M. (1969). *Frog, Where Are You?*. New York: Dial Press.
- Mayer, M. and Mayer, M. (1975). *One Frog Too Many*. New York: Dial Press.
- Menn, L., & Stoel-Gammon, C. (1995). Phonological development. In P. Fletcher and B. MacWhinney (Eds.), *The handbook of child language* (pp. 335-359). Oxford: Basil Blackwell.
- McElhinny, B. (1993). Copula and auxiliary contraction in the speech of white Americans. *American Speech*, 68, 371-399.

- Mitchell, B. (1985). *Old English Syntax*. Vols. I-II. Oxford: Clarendon Press.
- Mitchell, B., & Robinson, F. C. (2001). *A guide to Old English*. Oxford: Blackwell Publishers Ltd.
- Monaghan, J., & Ellis, A. (2002). Age of acquisition and the completeness of phonological representations. *Reading and Writing: An Interdisciplinary Journal*, 15, 759-788.
- Morrison, C., Chappell, T., & Ellis, A. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology*, 50A(3), 528-559.
- Munson, B. (2007). Lexical access, lexical representation, and vowel production. *Laboratory Phonology*, 9, 201-228.
- Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Speech, Language, and Hearing Research*, 47, 1048-1058.
- Newmeyer, F. J. (2006). On Gahl and Garnsey on grammar and usage. *Language*, 82, 399-404.
- Nooteboom, S., & Eefting, W. (1994). Evidence for the adaptive nature of speech on the phrase level and below. *Phonetica*, 51(1-3), 92-98.
- Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The journal of the Acoustical Society of America*, 54(5), 1235-1247.
- Palmer, F. (1986). *Mood and modality*. Cambridge: Cambridge University Press.
- Peramunage, D. Blumstein, S. E., Myers, E. B., Goldrick, M., & Baese-Berk, M. (2010). Phonological neighborhood effects in spoken word production: An fMRI study. *Journal of Cognitive Neuroscience*, 23, 593-603.
- Pick A. (1913). *Aphasia*. Transl./ed. J Brown, 1973. Springfield, IL: Thomas. (From German)
- Pickett, J. M. and Pollack, I. (1963). Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech*, 6, 151-164.
- Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee and P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137-157). Amsterdam: The John Benjamins Publishing Company.

- Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven, and N. Warner (Eds.), *Laboratory phonology VII (phonology and phonetics)* (pp. 101-140). Berlin: Mouton de Gruyter.
- Pierrehumbert, J. (2003). The voice of markedness. Paper presented at *Markedness and the lexicon: An IAP Workshop*. Cambridge, MA: MIT Press, January 24-25. (<http://babel.ling.northwestern.edu/~jbp/VoMarkcombine.pdf>).
- Pierrehumbert, J., & Gross, P. (2003). Community phonology. Paper presented at *the Chicago Linguistics Society*, Chicago, April 10-12. (<http://babel.ling.northwestern.edu/~jbp/cls03.pdf>).
- Pinker, S., Lebeaux, D., & Frost, L. A. (2008). Productivity constraints in the acquisition of the passive. *Cognition*, 26, 195-267.
- Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007) Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).
- Plank, F. (1985). The modal stories retold. *Studies in Language*, 8, 305-364.
- Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62(2-4), 146-159.
- Poplack, S. (1980). The notion of the plural in Puerto Rican Spanish: Competing constraints on /s/ deletion. In W. Labov, (Ed.), *Locating language in time and space* (pp. 55-67). Academic Press.
- Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar*. Oxford: Blackwell.
- Quené, H. (2007). On the just noticeable difference for tempo in speech. *Journal of Phonetics*, 35(3), 353-362.
- Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America*, 123(2), 1104-1113.
- Quirk, R., & Wren, C. L. (1957). *An Old English Grammar*. London: Methuen & Co Ltd.
- R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, retrieved 16 May, 2011 from <http://www.R-project.org/>

- Raymond, W. D., & Brown, E. L. (2012). Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. In S. Th. Gries and D. Divjak (Eds.), *Frequency effects in language learning and processing* (pp. 35-52). Berlin: De Gruyter Mouton.
- Redford, M. A. (2013). A comparative analysis of pausing in child and adult storytelling. *Applied Psycholinguistics*, 34, 569-589.
- Redford, M. A. (2014). The perceived clarity of children's speech varies as a function of their default articulation rate. *Journal of the Acoustical Society of America*, 135, 2952-2963.
- Redford, M. A. (Forthcoming). The acquisition of temporal patterns. In M. A. Redford (Ed.), *The Handbook of Speech Production*. Hoboken, NJ: John Wiley & Sons.
- Revelle, W. (2014). psych: Procedures for psychological, psychometric, and personality research. Version 1.4.8.11. Package for R. <http://cran.r-project.org/web/packages/psych/index.html>
- Rondal, J. A. (1980). Fathers' and mothers' speech in early language development. *Journal of Child Language*, 7(2), 353-369.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical cues in language acquisition: Word segmentation by infants. *COGSCI-96*, 376-380.
- Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, 60, 362-267.
- Schachter, S., Rauscher, F., Christenfeld, N., & Crone, K. T. (1994). The vocabularies of academia. *Psychological Science*, 5, 37-41.
- Scherre, M. M. P., & Naro, A. J. 1991. Marking in discourse: "Birds of a feather". *Language Variation and Change*, 3(1), 23-32.
- Schneider, U. (2014) *Frequency, chunks and hesitations: A usage-based analysis of chunking in English*. Freiburg: PhD. Dissertation. Albert-Ludwigs-Universität Freiburg.
- Schnadt, M. J. (2009). Lexical influences on disfluency production. Edinburgh: PhD. Dissertation. University of Edinburgh.
- Schuchardt, H. (1885). *Über die Lautgesetze: Gegen die Junggrammatiker*. Berlin: R. Oppenheim.
- Schürmann, T. (2004). Bias analysis in entropy estimation. *Journal of Physics A: Mathematical and General*, 37(27), L295-L301.

- Schürmann, T., & Grassberger, P. (1996). Entropy estimation of symbol sequences. *Chaos*, 6(3), 414-427.
- Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133, 140-155.
- Shannon, C. E. (1948). The mathematics theory of communication. *Bell System Technical Journal*, 27, 379-423.
- Shih, S., & Grafmiller, J. (2011). Weighing in on end weight. In Paper presented at the 2011 Annual meeting of the Linguistic Society of America.
- Shockey, L. (1998). Perception of reduced forms by non-native speakers of English. In D. Duez (Ed.), *Sound patterns of spontaneous speech* (pp. 97-100). Aix: ESCA.
- Shockey, L. (2003). *Sound patterns of spoken English*. Cambridge: Blackwell.
- Sirsa, H., & Redford, M. A. (2011). Towards understanding the protracted acquisition of English rhythm. In *Proceedings from the 17th International Congress of Phonetic Sciences* (pp. 1862-1865).
- Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of Communication Disorders*, 39, 331-349.
- Smith, B. L. (1978). Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics*, 6, 37-67.
- Smith, B. L., & Kenney, M. K. (1994). Variability control in speech production tasks performed by adults and children. *The Journal of the Acoustical Society of America*, 96, 699-705.
- Smith, B. L., Kenney, M. K., & Hussain, S. (1996). A longitudinal investigation of duration and temporal variability in children's speech production. *The Journal of the Acoustical Society of America*, 99, 2344-2349.
- Snow, C. E. (1995). Issues in the study of input: Finetuning, universality, individual and developmental differences, and necessary causes. In P. Fletcher and B. MacWhinney (Eds.), *The handbook of child language* (pp. 180-193). Oxford: Blackwell.
- Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech, Language, and Hearing Research*, 37(4), 831-840.
- Stadthagen-Gonzalez, H., & Davis, C. (2006). The Bristol norms for age of acquisition, imageability and familiarity. *Behavior Research Methods*, 38(4), 598-605.
- Stephany, U. (1986). Modality. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 375-400). Cambridge: Cambridge University Press.

- Strobl, C., Boulesteix, A., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*, 8(25). <http://www.biomedcentral.com/1471-2105/8/25>.
- Strobl, C., Boulesteix, A., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(307). <http://www.biomedcentral.com/1471-2105/9/307>.
- Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14(4), 323-348.
- Stromswold K., Caplan, D. Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by Positron Emission Tomography. *Brain and Language*, 52, 452-473.
- Swanson, L., L. Leonard, L. and J. Gandour. (1992). Vowel duration in mothers' speech to young children. *Journal of Speech, Language, and Hearing Research*, 35(3), 617-625.
- Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(02), 135-178.
- Tannenbaum, P. H., Williams, F., & Hillier, C. S. (1965). Word predictability in the environments of hesitations. *Journal of Verbal Learning and Verbal Behavior*, 4(2), 134-140.
- Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, 1(2), 147-165.
- Tomasello, M. (2008). One child's early talk about possession. In J. Newman (Ed.), *The linguistics of giving* (pp. 349-373). Amsterdam: John Benjamins.
- Tomasello, M. (2003). *Constructing a Language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Torres Cacoullos, R. (1999). Construction frequency and reductive change: Diachronic and register variation in Spanish clitic climbing. *Language Variation and Change*, 11(2), 143-170.
- Torres Cacoullos, R., & Travis, C. E. (2013). Prosody, priming and particular constructions: The patterning of English first-person singular subject expression in conversation. *Journal of Pragmatics*, 63, 19-34.

- Cacoullos, R. T., & Walker, J. A. (2011). Collocations in grammaticalization and variation. In B. Heine & H. Narrog (Eds.) *The Oxford handbook of grammaticalization* (pp. 225-238). Oxford: Oxford University Press.
- Traugott, E. C. (2003). Constructions in grammaticalization. In B. D. Joseph & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 624-647). Oxford: Blackwell.
- Traugott, E. C. (2011). Grammaticalization and mechanisms of change. In B. Heine & H. Narrog, (Eds.), *The Oxford handbook of grammaticalization* (pp. 19-31). Oxford: Oxford University Press.
- Travis, C. E. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, 19, 101-35.
- Tucker, B. V. (2011). The effect of reduction on the processing of flaps and/g/in isolated words. *Journal of Phonetics*, 39(3), 312-318.
- Ullman, M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Neuroscience*, 2, 717-727.
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 231-270.
- van Bergem, D. (1995). *Acoustic and lexical vowel reduction*. Amsterdam: IFOTT.
- van de Ven, M., Tucker, B. V., & Ernestus, M. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory and Cognition*, 39, 1301-1316.
- van der Gaaf, W. (1931). *Beon* and *habban* connected with an inflected infinitive. *English Studies*, 13, 176-188.
- van Son, R. J. J. H., & Pols, L. C. W. (2003). Information structure and efficiency in speech production. In Paper presented at the 2003 Eurospeech conference, Geneva, Switzerland.
- Visser, F. Th. (1966). *An historical syntax of the English language*. Vol. II. Leiden: E. J. Brill.
- Visser, F. Th. (1969). *An historical syntax of the English language*. Vol. IIIa. Leiden: E. J. Brill.
- Visser, F. Th. (1973). *An historical syntax of the English language*. Vol. IIIb. Leiden: E. J. Brill.

- Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 735.
- Warner, N., Fountain, A., & Tucker, B. V. (2009). Cues to perception of reduced flaps. *The Journal of the Acoustical Society of America*, 125(5), 3317-3327.
- Wasow, T., Jaeger, T. F., & Orr, D. (2006). Lexical variation in relativizer frequency. In H. Wiese and H. Simon (Eds.), *Expecting the Unexpected: Exceptions in Grammar* (pp. 175-196). Berlin: Walter de Gruyter.
- Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, 106, 156–157.
- Watson, P. J., & Munson, B. (2007). A comparison of vowel acoustics between older and younger adults. In *Proceedings of the 16th international congress of phonetic sciences* (pp. 561–564).
- White, L. (2002). *English speech timing: a domain and locus approach*. Edinburgh: PhD dissertation. University of Edinburgh.
- Wolfram, W. (1969). *A sociolinguistic description of Detroit Negro speech*. Washington, DC: Center for Applied Linguistics.
- Wood, S. (2014). mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation. Package for R. <http://cran.r-project.org/web/packages/mgcv/index.html>
- Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden & R. Temple (Eds.), *Papers in laboratory phonology VI* (pp. 26–50). Cambridge: Cambridge University Press.
- Xiao, R., McEnery, T., & Qian, Y. (2006). Passive constructions in English and Chinese: a corpus-based contrastive study. *Languages in Contrast*, 6, 109-149.
- Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. Paper presented at the *International Conference on Spoken Language Processing (Interspeech 2006)*. Pittsburgh.
- Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard studies in classical philology*, 15, 1-95.
- Zipf, G. K. (1935). *The psycho-biology of language*. Boston, MA: Houghton Mifflin
- Zwaardemaker, H., & Eijkman, L. (1928). *Leerboek der Phonetiek: inzonderheid met betrekking tot het Standaard-Nederlandsch*. Haarlem: Erven F. Bohn.