

Investigating variable linguistic systems in flux in 17<sup>th</sup> century Massachusetts English

Samuel Gardiner

A thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the  
Doctorate in Philosophy degree in Linguistics

Department of Linguistics  
Faculty of Arts  
Université d'Ottawa/University of Ottawa

# TABLE OF CONTENTS

|                                                                                  |           |
|----------------------------------------------------------------------------------|-----------|
| <b>LIST OF TABLES.....</b>                                                       | <b>VI</b> |
| <b>LIST OF FIGURES.....</b>                                                      | <b>IX</b> |
| <b>ABSTRACT .....</b>                                                            | <b>X</b>  |
| <b>ACKNOWLEDGEMENTS .....</b>                                                    | <b>XI</b> |
| <b>CHAPTER 1: INTRODUCTION.....</b>                                              | <b>1</b>  |
| 1.1 THE 17TH CENTURY AS A LINGUISTIC PERIOD .....                                | 1         |
| 1.2 EARLY AMERICAN ENGLISH.....                                                  | 2         |
| 1.3 DATA IN DIACHRONIC RESEARCH .....                                            | 2         |
| 1.4 PUTTING THE <i>SALEM WITCHCRAFT PAPERS</i> INTO HISTORICAL CONTEXT.....      | 3         |
| 1.5 LINGUISTIC VARIABLES UNDER INVESTIGATION.....                                | 5         |
| 1.5.1 <i>Relativisation</i> .....                                                | 5         |
| 1.5.2 <i>Do-Support</i> .....                                                    | 6         |
| 1.6 THE VARIATIONIST APPROACH .....                                              | 7         |
| 1.7 MAJOR RESEARCH OBJECTIVES .....                                              | 9         |
| <b>CHAPTER 2: CRITICAL EVALUATION OF THE DATA AND THEORETICAL FRAMEWORK.....</b> | <b>11</b> |
| 2.1 THE SALEM WITCH TRIALS .....                                                 | 11        |
| 2.2 THE <i>SALEM WITCHCRAFT PAPERS</i> .....                                     | 13        |
| 2.3 RECONSTRUCTING ORALITY .....                                                 | 15        |
| 2.4 THEORETICAL FRAMEWORK .....                                                  | 20        |
| <b>CHAPTER 3: RELATIVISATION.....</b>                                            | <b>22</b> |
| 3.1 INTRODUCTION TO RELATIVISATION .....                                         | 22        |
| 3.2 HISTORY OF RELATIVISATION IN ENGLISH.....                                    | 22        |
| 3.2.1 <i>Key Works</i> .....                                                     | 22        |
| 3.2.2 <i>Relativisation in Present Day English</i> .....                         | 26        |
| 3.2.3 <i>Relativisation in Old English</i> .....                                 | 32        |
| 3.2.4 <i>Relativisation in Middle English</i> .....                              | 35        |
| 3.2.5 <i>Relativisation in Early Modern English</i> .....                        | 39        |
| 3.3 METHODOLOGY .....                                                            | 45        |
| 3.3.1 <i>Variable Context</i> .....                                              | 45        |
| 3.3.2 <i>Internal Factor Groups</i> .....                                        | 47        |
| 3.3.2.1 Syntactic Function of the Relative Marker .....                          | 47        |
| 3.3.2.2 Animacy .....                                                            | 48        |
| 3.3.2.3 Matrix Clause Type .....                                                 | 48        |
| 3.3.2.4 Processing Constraints.....                                              | 49        |
| 3.3.2.5 Antecedent Type and Definiteness.....                                    | 51        |
| 3.3.3 <i>External Factor Groups</i> .....                                        | 52        |
| 3.3.3.1 Speaker Sex.....                                                         | 52        |
| 3.3.3.2 Speaker Birthplace.....                                                  | 53        |
| 3.3.3.3 Speaker Age .....                                                        | 53        |
| 3.3.3.4 Speaker Social Status .....                                              | 53        |
| 3.3.3.5 Document Type .....                                                      | 55        |
| 3.4 RESULTS .....                                                                | 55        |
| 3.4.1 <i>Overall Distribution</i> .....                                          | 55        |

|                                                                                                                                                                                                                                                                                |            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 3.4.1.1 Set Phrases .....                                                                                                                                                                                                                                                      | 56         |
| 3.4.2 Syntactic Function of the Relative Marker.....                                                                                                                                                                                                                           | 59         |
| 3.4.3 Subject Relativisers.....                                                                                                                                                                                                                                                | 61         |
| 3.4.3.1 Animacy .....                                                                                                                                                                                                                                                          | 61         |
| 3.4.3.2 Matrix Clause Type .....                                                                                                                                                                                                                                               | 63         |
| 3.4.3.3 Processing Constraints.....                                                                                                                                                                                                                                            | 64         |
| 3.4.3.4 Antecedent Type.....                                                                                                                                                                                                                                                   | 66         |
| 3.4.3.5 Sex .....                                                                                                                                                                                                                                                              | 68         |
| 3.4.3.6 Birthplace.....                                                                                                                                                                                                                                                        | 69         |
| 3.4.3.7 Age.....                                                                                                                                                                                                                                                               | 70         |
| 3.4.3.8 Social Status.....                                                                                                                                                                                                                                                     | 72         |
| 3.4.3.9 Document Type .....                                                                                                                                                                                                                                                    | 74         |
| 3.4.4 Non-Subject Relativisers .....                                                                                                                                                                                                                                           | 74         |
| 3.4.4.1 Animacy .....                                                                                                                                                                                                                                                          | 74         |
| 3.4.4.2 Matrix Clause Type .....                                                                                                                                                                                                                                               | 76         |
| 3.4.4.3 Processing Constraints.....                                                                                                                                                                                                                                            | 76         |
| 3.4.4.4 Preposition Stranding .....                                                                                                                                                                                                                                            | 78         |
| 3.4.4.5 Antecedent Type.....                                                                                                                                                                                                                                                   | 80         |
| 3.4.4.6 Sex .....                                                                                                                                                                                                                                                              | 81         |
| 3.4.4.7 Birthplace.....                                                                                                                                                                                                                                                        | 82         |
| 3.4.4.8 Age.....                                                                                                                                                                                                                                                               | 82         |
| 3.4.4.9 Social Status.....                                                                                                                                                                                                                                                     | 83         |
| 3.4.4.10 Document Type .....                                                                                                                                                                                                                                                   | 85         |
| 3.4.5 Statistical Analysis .....                                                                                                                                                                                                                                               | 86         |
| 3.4.5.1 Subject Relatives.....                                                                                                                                                                                                                                                 | 87         |
| 3.4.5.2 Non-Subject Relatives.....                                                                                                                                                                                                                                             | 91         |
| 3.5 DISCUSSION .....                                                                                                                                                                                                                                                           | 97         |
| 3.5.1 Objective (i): to reconstruct the variable system of the target linguistic phenomenon as it exists in contemporaneous speech-related data.....                                                                                                                           | 97         |
| 3.5.2 Objective (ii): to refute, confirm, or refine existing claims in the literature regarding the target phenomenon ....                                                                                                                                                     | 99         |
| 3.5.2.1 Confirmation of previous claims .....                                                                                                                                                                                                                                  | 99         |
| 3.5.2.2 Refinement of previous claims .....                                                                                                                                                                                                                                    | 100        |
| 3.5.2.3 Refutation of previous claims .....                                                                                                                                                                                                                                    | 101        |
| 3.5.3 Objective (iii): to demonstrate the value of the Salem Witchcraft Papers as a resource for investigating linguistic variation in 17 <sup>th</sup> century Massachusetts English (and more broadly, early colonial American English) in its socio-historical context..... | 101        |
| <b>CHAPTER 4: DO-SUPPORT.....</b>                                                                                                                                                                                                                                              | <b>103</b> |
| 4.1 INTRODUCTION TO <i>Do</i> -SUPPORT .....                                                                                                                                                                                                                                   | 103        |
| 4.2 REVIEW OF THE LITERATURE ON <i>Do</i> -SUPPORT .....                                                                                                                                                                                                                       | 104        |
| 4.2.1 <i>Do</i> -Support in Present Day English .....                                                                                                                                                                                                                          | 106        |
| 4.2.2 <i>The Origins of Do-Support</i> .....                                                                                                                                                                                                                                   | 108        |
| 4.2.2.1 Internally-Motivated Claims .....                                                                                                                                                                                                                                      | 108        |
| 4.2.2.2 Externally-Motivated Claims.....                                                                                                                                                                                                                                       | 110        |
| 4.2.3 <i>History of Do-Support</i> .....                                                                                                                                                                                                                                       | 113        |
| 4.2.3.1 <i>Do</i> -Support in Old English.....                                                                                                                                                                                                                                 | 114        |
| 4.2.3.2 <i>Do</i> -Support in Middle English .....                                                                                                                                                                                                                             | 114        |
| 4.2.3.3 <i>Do</i> -Support in Early Modern English.....                                                                                                                                                                                                                        | 116        |
| 4.2.4 <i>Internal Factors Affecting the Selection of Do-Support in Early Modern English</i> .....                                                                                                                                                                              | 123        |
| 4.2.4.1 Verb Morphology.....                                                                                                                                                                                                                                                   | 123        |
| 4.2.4.2 Tense .....                                                                                                                                                                                                                                                            | 124        |
| 4.2.4.3 Aspect .....                                                                                                                                                                                                                                                           | 124        |

|                                                                                                   |     |
|---------------------------------------------------------------------------------------------------|-----|
| 4.2.4.4 Adverbs.....                                                                              | 124 |
| 4.2.4.5 Clause Structure.....                                                                     | 125 |
| 4.2.4.6 Lexical Effects .....                                                                     | 125 |
| 4.2.5 <i>External Factors Affecting the Selection of Do-Support in Early Modern English</i> ..... | 126 |
| 4.2.5.1 Geographical Origin .....                                                                 | 126 |
| 4.2.5.2 Age.....                                                                                  | 127 |
| 4.2.5.3 Sex .....                                                                                 | 127 |
| 4.2.5.4 Document Type .....                                                                       | 127 |
| 4.2.6 <i>The Salem Witchcraft Papers as a Window on the Development of Do-Support</i> .....       | 128 |
| 4.3 METHODOLOGY .....                                                                             | 129 |
| 4.3.1 <i>Variable Context</i> .....                                                               | 129 |
| 4.3.2 <i>Internal Factor Groups</i> .....                                                         | 132 |
| 4.3.2.1 Sentence Type .....                                                                       | 132 |
| 4.3.2.2 Clause Type .....                                                                         | 132 |
| 4.3.2.3 Main Verb Morphology .....                                                                | 133 |
| 4.3.2.4 Tense .....                                                                               | 135 |
| 4.3.2.5 Aspect .....                                                                              | 135 |
| 4.3.2.6 Adverbs.....                                                                              | 136 |
| 4.3.2.7 Adverb Position .....                                                                     | 137 |
| 4.3.3 <i>Lexical Collocations</i> .....                                                           | 137 |
| 4.3.4 <i>External Factor Groups</i> .....                                                         | 138 |
| 4.3.4.1 Speaker Sex.....                                                                          | 138 |
| 4.3.4.2 Speaker Age .....                                                                         | 138 |
| 4.3.4.3 Speaker Birthplace.....                                                                   | 139 |
| 4.3.4.4 Speaker Social Status.....                                                                | 139 |
| 4.3.4.5 Text Type .....                                                                           | 139 |
| 4.4 RESULTS .....                                                                                 | 140 |
| 4.4.1 <i>Overall Distribution</i> .....                                                           | 140 |
| 4.4.1.1. Set Phrases .....                                                                        | 141 |
| 4.4.2 <i>Internal Factors</i> .....                                                               | 142 |
| 4.4.2.1 Sentence Type and Polarity .....                                                          | 142 |
| 4.4.2.2 Verb Morphology.....                                                                      | 144 |
| 4.4.2.3 Tense and Aspect.....                                                                     | 146 |
| 4.4.2.4 Clause Type.....                                                                          | 149 |
| 4.4.2.5 Adverbs.....                                                                              | 150 |
| 4.4.3 <i>Lexical Effects</i> .....                                                                | 155 |
| 4.4.3.1 The <i>Know-Verbs</i> .....                                                               | 157 |
| 4.4.3.2 Verbs Expressing Mental Processes.....                                                    | 163 |
| 4.4.3.3 Verbs Expressing Speech Acts.....                                                         | 166 |
| 4.4.3.4 Emotive Verbs.....                                                                        | 169 |
| 4.4.3.5 Legal Terminology.....                                                                    | 171 |
| 4.4.3.6 Other High-Frequency Verbs .....                                                          | 174 |
| 4.4.3.7 Germanic vs. Latinate Verbs .....                                                         | 178 |
| 4.4.3.8 Summary of Lexical Effects .....                                                          | 179 |
| 4.4.4 <i>External Factors</i> .....                                                               | 181 |
| 4.4.4.1 Sex .....                                                                                 | 181 |
| 4.4.4.2 Age.....                                                                                  | 183 |
| 4.4.4.3 Birthplace.....                                                                           | 185 |
| 4.4.4.4 Social Status.....                                                                        | 187 |
| 4.4.4.5 Document Type .....                                                                       | 189 |
| 4.4.5 <i>Statistical Analysis</i> .....                                                           | 191 |
| 4.5 DISCUSSION .....                                                                              | 196 |

|                                                                                                                                                                                                                                                                                |            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 4.5.1 Objective (i): to reconstruct the variable system of the target linguistic phenomenon as it exists in contemporaneous speech-related data.....                                                                                                                           | 197        |
| 4.5.2 Objective (ii): to refute, confirm, or refine existing claims in the literature regarding the target phenomenon..                                                                                                                                                        | 200        |
| 4.5.2.1 Confirmation of previous claims.....                                                                                                                                                                                                                                   | 200        |
| 4.5.2.2 Refinement of previous claims.....                                                                                                                                                                                                                                     | 200        |
| 4.5.2.3 Refutation of previous claims.....                                                                                                                                                                                                                                     | 203        |
| 4.5.3 Objective (iii): to demonstrate the value of the Salem Witchcraft Papers as a resource for investigating linguistic variation in 17 <sup>th</sup> century Massachusetts English (and more broadly, early colonial American English) in its socio-historical context..... | 204        |
| <b>CHAPTER 5: CONTEXTUALIZING QUANTITATIVE PATTERNS .....</b>                                                                                                                                                                                                                  | <b>207</b> |
| 5.1 SUMMARY OF MAJOR FINDINGS.....                                                                                                                                                                                                                                             | 208        |
| 5.1.1 Relativisation.....                                                                                                                                                                                                                                                      | 208        |
| 5.1.2 Do-Support .....                                                                                                                                                                                                                                                         | 210        |
| 5.2 THEORETICAL FRAMEWORK .....                                                                                                                                                                                                                                                | 210        |
| 5.3 THE SALEM WITCHCRAFT PAPERS.....                                                                                                                                                                                                                                           | 212        |
| 5.4 THE TARGET VARIABLES.....                                                                                                                                                                                                                                                  | 213        |
| 5.5 FUTURE AVENUES FOR INVESTIGATION .....                                                                                                                                                                                                                                     | 214        |
| <b>BIBLIOGRAPHY .....</b>                                                                                                                                                                                                                                                      | <b>216</b> |

## LIST OF TABLES

|                                                                                                                                                                              |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| TABLE 1: OVERALL DISTRIBUTION OF VARIANTS .....                                                                                                                              | 56 |
| TABLE 2: OVERALL DISTRIBUTION OF VARIANTS IN SET PHRASES.....                                                                                                                | 57 |
| TABLE 3: DISTRIBUTION OF SET PHRASES .....                                                                                                                                   | 58 |
| TABLE 4: CONTRIBUTION OF SYNTACTIC FUNCTION OF THE RELATIVE MARKER TO THE SELECTION OF VARIANTS .....                                                                        | 60 |
| TABLE 5: CONTRIBUTION OF ANIMACY OF THE ANTECEDENT TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES.....                                                                    | 61 |
| TABLE 6: CONTRIBUTION OF MATRIX CLAUSE CONSTRUCTION TYPE TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES.....                                                              | 63 |
| TABLE 7: CONTRIBUTION OF LENGTH OF THE RELATIVE CLAUSE TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES .....                                                               | 64 |
| TABLE 8: CONTRIBUTION OF ADJACENCY TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES .....                                                                                   | 65 |
| TABLE 9: CONTRIBUTION OF ANTECEDENT TYPE TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES .....                                                                             | 67 |
| TABLE 10: CONTRIBUTION OF DEFINITENESS OF THE ANTECEDENT TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES .....                                                             | 67 |
| TABLE 11: CONTRIBUTION OF SPEAKER SEX TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES .....                                                                                | 69 |
| TABLE 12: CONTRIBUTION OF SPEAKER BIRTHPLACE TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES .....                                                                         | 69 |
| TABLE 13: CONTRIBUTION OF SPEAKER AGE TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES.....                                                                                 | 70 |
| TABLE 14: CONTRIBUTION OF SPEAKER AGE TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES (WHO AND WHICH SEPARATED) .....                                                      | 71 |
| TABLE 15: CONTRIBUTION OF SPEAKER SOCIAL STATUS TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES .....                                                                      | 72 |
| TABLE 16: CONTRIBUTION OF SPEAKER ROLE TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES .....                                                                               | 73 |
| TABLE 17: CONTRIBUTION OF DOCUMENT TYPE TO THE SELECTION OF VARIANTS IN SUBJECT RELATIVES.....                                                                               | 74 |
| TABLE 18: CONTRIBUTION OF ANIMACY OF THE ANTECEDENT TO THE SELECTION OF VARIANTS IN NON-SUBJECT RELATIVES .....                                                              | 75 |
| TABLE 19: CONTRIBUTION OF MATRIX CLAUSE CONSTRUCTION TYPE TO THE SELECTION OF VARIANTS IN NON-SUBJECT RELATIVES .....                                                        | 76 |
| TABLE 20: CONTRIBUTION OF LENGTH OF THE RELATIVE CLAUSE TO THE SELECTION OF VARIANTS IN NON-SUBJECT RELATIVES.....                                                           | 77 |
| TABLE 21: CONTRIBUTION OF ADJACENCY TO THE SELECTION OF VARIANTS IN NON-SUBJECT RELATIVES.....                                                                               | 77 |
| TABLE 22: CONTRIBUTION OF PREPOSITION STRANDING TO THE SELECTION OF VARIANTS .....                                                                                           | 78 |
| TABLE 23: CONTRIBUTION OF GRAMMATICAL CATEGORY OF THE ANTECEDENT TO THE SELECTION OF VARIANTS IN NON-SUBJECT RELATIVES .....                                                 | 80 |
| TABLE 24: CONTRIBUTION OF DEFINITENESS OF THE ANTECEDENT TO THE SELECTION OF VARIANTS IN NON-SUBJECT RELATIVES.....                                                          | 81 |
| TABLE 25: CONTRIBUTION OF SPEAKER SEX TO THE SELECTION OF VARIANTS IN NON-SUBJECT RELATIVES .....                                                                            | 81 |
| TABLE 26: CONTRIBUTION OF SPEAKER BIRTHPLACE TO THE SELECTION OF VARIANTS IN NON-SUBJECT RELATIVES .....                                                                     | 82 |
| TABLE 27: CONTRIBUTION OF SPEAKER AGE TO THE SELECTION OF VARIANTS IN NON-SUBJECT RELATIVES .....                                                                            | 83 |
| TABLE 28: CONTRIBUTION OF SPEAKER SOCIAL STATUS TO THE SELECTION OF VARIANTS IN NON-SUBJECT RELATIVES .....                                                                  | 84 |
| TABLE 29: CONTRIBUTION OF SPEAKER ROLE TO THE SELECTION OF VARIANTS IN NON-SUBJECT RELATIVES .....                                                                           | 85 |
| TABLE 30: CONTRIBUTION OF DOCUMENT TYPE TO THE SELECTION OF VARIANTS IN NON-SUBJECT RELATIVES .....                                                                          | 86 |
| TABLE 31: RBRUL MIXED-EFFECTS LOGISTIC REGRESSION ANALYSIS OF THE CONTRIBUTION OF LINGUISTIC FACTORS TO THE SELECTION OF RELATIVISER<br>THAT IN SUBJECT RELATIVES .....      | 88 |
| TABLE 32: RBRUL MIXED-EFFECTS LOGISTIC REGRESSION ANALYSIS OF THE CONTRIBUTION OF SOCIAL FACTORS TO THE SELECTION OF RELATIVISER THAT<br>IN SUBJECT RELATIVES.....           | 89 |
| TABLE 33: RBRUL MIXED-EFFECTS LOGISTIC REGRESSION ANALYSIS OF THE CONTRIBUTION OF LINGUISTIC FACTORS TO THE SELECTION OF RELATIVISER<br>WHICH IN SUBJECT RELATIVES.....      | 90 |
| TABLE 34: RBRUL MIXED-EFFECTS LOGISTIC REGRESSION ANALYSIS OF THE CONTRIBUTION OF SOCIAL FACTORS TO THE SELECTION OF RELATIVISER<br>WHICH IN SUBJECT RELATIVES.....          | 91 |
| TABLE 35: RBRUL MIXED-EFFECTS LOGISTIC REGRESSION ANALYSIS OF THE CONTRIBUTION OF LINGUISTIC FACTORS TO THE SELECTION OF RELATIVISER<br>THAT IN NON-SUBJECT RELATIVES.....   | 92 |
| TABLE 36: RBRUL MIXED-EFFECTS LOGISTIC REGRESSION ANALYSIS OF THE CONTRIBUTION OF SOCIAL FACTORS TO THE SELECTION OF RELATIVISER THAT<br>IN NON-SUBJECT RELATIVES .....      | 93 |
| TABLE 37: RBRUL MIXED-EFFECTS LOGISTIC REGRESSION ANALYSIS OF THE CONTRIBUTION OF LINGUISTIC FACTORS TO THE SELECTION OF RELATIVISER<br>WHICH IN NON-SUBJECT RELATIVES ..... | 94 |
| TABLE 38: RBRUL MIXED-EFFECTS LOGISTIC REGRESSION ANALYSIS OF THE CONTRIBUTION OF SOCIAL FACTORS TO THE SELECTION OF RELATIVISER<br>WHICH IN NON-SUBJECT RELATIVES .....     | 95 |

|                                                                                                                                                                             |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| TABLE 39: RBRUL MIXED-EFFECTS LOGISTIC REGRESSION ANALYSIS OF THE CONTRIBUTION OF LINGUISTIC FACTORS TO THE SELECTION OF THE ZERO<br>RELATIVE IN NON-SUBJECT RELATIVES..... | 96  |
| TABLE 40: RBRUL MIXED-EFFECTS LOGISTIC REGRESSION ANALYSIS OF THE CONTRIBUTION OF SOCIAL FACTORS TO THE SELECTION OF THE ZERO<br>RELATIVE IN NON-SUBJECT RELATIVES.....     | 97  |
| TABLE 41: OVERALL DISTRIBUTION OF VARIANTS.....                                                                                                                             | 141 |
| TABLE 42: DISTRIBUTION OF <i>DO</i> -SUPPORT IN SET PHRASES.....                                                                                                            | 141 |
| TABLE 43: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                                         | 142 |
| TABLE 44: CONTRIBUTION OF POLARITY TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                                              | 142 |
| TABLE 45: CONTRIBUTION OF SENTENCE TYPE AND POLARITY TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                            | 144 |
| TABLE 46: CONTRIBUTION OF VERB STRENGTH TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                                         | 145 |
| TABLE 47: CONTRIBUTION OF VERB STRENGTH TO THE SELECTION OF <i>DO</i> -SUPPORT IN AFFIRMATIVE DECLARATIVES.....                                                             | 145 |
| TABLE 48: CONTRIBUTION OF TENSE TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                                                 | 146 |
| TABLE 49: CONTRIBUTION OF TENSE AND SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                               | 147 |
| TABLE 50: CONTRIBUTION OF HABITUAL ASPECT TO THE SELECTION OF <i>DO</i> -SUPPORT IN PAST TENSE CONTEXTS.....                                                                | 148 |
| TABLE 51: CONTRIBUTION OF ASPECT AND SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                              | 148 |
| TABLE 52: CONTRIBUTION OF CLAUSE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                                           | 149 |
| TABLE 53: CONTRIBUTION OF CLAUSE TYPE AND SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                         | 150 |
| TABLE 54: CONTRIBUTION OF ADVERBS TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                                               | 151 |
| TABLE 55: CONTRIBUTION OF ADVERBS AND SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                             | 152 |
| TABLE 56: CONTRIBUTION OF ADVERB POSITION TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                                       | 153 |
| TABLE 57: DISTRIBUTION OF <i>DO</i> -SUPPORT BY ADVERB PLACEMENT.....                                                                                                       | 153 |
| TABLE 58: CONTRIBUTION OF ADVERB PLACEMENT AND SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                    | 155 |
| TABLE 59: DISTRIBUTION OF <i>DO</i> -SUPPORT WITH HIGH-FREQUENCY VERBS (100+ TOKENS).....                                                                                   | 156 |
| TABLE 60: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB KNOW.....                                                                      | 158 |
| TABLE 61: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB COME.....                                                                      | 159 |
| TABLE 62: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH MAIN VERB <i>DO</i> .....                                                               | 159 |
| TABLE 63: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB SEE.....                                                                       | 160 |
| TABLE 64: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB GO.....                                                                        | 161 |
| TABLE 65: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB SAY.....                                                                       | 162 |
| TABLE 66: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB THINK.....                                                                     | 163 |
| TABLE 67: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB HEAR.....                                                                      | 164 |
| TABLE 68: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB BELIEVE.....                                                                   | 165 |
| TABLE 69: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB LOOK.....                                                                      | 165 |
| TABLE 70: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB TELL.....                                                                      | 166 |
| TABLE 71: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB ASK.....                                                                       | 167 |
| TABLE 72: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB ANSWER.....                                                                    | 168 |
| TABLE 73: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB CRY.....                                                                       | 168 |
| TABLE 74: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB AFFLICT.....                                                                   | 170 |
| TABLE 75: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB HURT.....                                                                      | 170 |
| TABLE 76: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB TORMENT.....                                                                   | 171 |
| TABLE 77: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB TESTIFY.....                                                                   | 172 |
| TABLE 78: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB OWN.....                                                                       | 172 |
| TABLE 79: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB PRESENT.....                                                                   | 173 |
| TABLE 80: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB AFFIRM.....                                                                    | 173 |
| TABLE 81: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB FALL.....                                                                      | 174 |
| TABLE 82: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB HAVE.....                                                                      | 175 |
| TABLE 83: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB MAKE.....                                                                      | 175 |
| TABLE 84: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB APPEAR.....                                                                    | 176 |
| TABLE 85: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB BRING.....                                                                     | 177 |

|                                                                                                                                                        |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| TABLE 86: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB GIVE .....                                                | 177 |
| TABLE 87: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB TAKE.....                                                 | 178 |
| TABLE 88: CONTRIBUTION OF SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT WITH THE VERB STAND .....                                               | 178 |
| TABLE 89: CONTRIBUTION OF SENTENCE TYPE AND VERB LANGUAGE OF ORIGIN TO THE SELECTION OF <i>DO</i> -SUPPORT .....                                       | 179 |
| TABLE 90: SUMMARY OF LEXICAL EFFECTS ON THE SELECTION OF <i>DO</i> -SUPPORT.....                                                                       | 181 |
| TABLE 91: CONTRIBUTION OF SPEAKER SEX TO THE SELECTION OF <i>DO</i> -SUPPORT .....                                                                     | 181 |
| TABLE 92: CONTRIBUTION OF SPEAKER SEX AND SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT .....                                                   | 183 |
| TABLE 93: CONTRIBUTION OF SPEAKER AGE TO THE SELECTION OF <i>DO</i> -SUPPORT .....                                                                     | 183 |
| TABLE 94: CONTRIBUTION OF SPEAKER AGE AND SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT .....                                                   | 185 |
| TABLE 95: CONTRIBUTION OF SPEAKER AGE, SPEAKER SEX, AND SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT .....                                     | 185 |
| TABLE 96: CONTRIBUTION OF SPEAKER BIRTHPLACE TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                               | 186 |
| TABLE 97: CONTRIBUTION OF SPEAKER BIRTHPLACE AND SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT .....                                            | 187 |
| TABLE 98: CONTRIBUTION OF SPEAKER SOCIAL STATUS TO THE SELECTION OF <i>DO</i> -SUPPORT.....                                                            | 188 |
| TABLE 99: CONTRIBUTION OF SPEAKER SOCIAL STATUS AND SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT .....                                         | 189 |
| TABLE 100: CONTRIBUTION OF DOCUMENT TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT .....                                                                  | 190 |
| TABLE 101: CONTRIBUTION OF DOCUMENT TYPE AND SENTENCE TYPE TO THE SELECTION OF <i>DO</i> -SUPPORT .....                                                | 191 |
| TABLE 102: RBRUL MIXED-EFFECTS LOGISTIC REGRESSION ANALYSIS OF THE CONTRIBUTION OF LINGUISTIC FACTORS TO THE SELECTION OF <i>DO</i> -<br>SUPPORT ..... | 192 |
| TABLE 103: RBRUL MIXED-EFFECTS LOGISTIC REGRESSION ANALYSIS OF THE CONTRIBUTION OF SOCIAL FACTORS TO THE SELECTION OF <i>DO</i> -SUPPORT<br>.....      | 195 |



## LIST OF FIGURES

|                                                                                                                       |     |
|-----------------------------------------------------------------------------------------------------------------------|-----|
| FIGURE 1: THE DISTRIBUTION OF <i>DO</i> -SUPPORT FROM 1400-1700 (FROM LOS, 2015: 111, BASED ON ELLEGÅRD, 1953). ..... | 117 |
|-----------------------------------------------------------------------------------------------------------------------|-----|

## ABSTRACT

This thesis reconstructs the variable systems of two linguistic phenomena undergoing change in 17<sup>th</sup> century English – restrictive relativisation (Ball, 1996; Romaine, 1982), and *do*-support (Ellegård, 1953; Söderlund, 2017). Innovative here is the use of a digitized compendium of the *Salem Witchcraft Papers* (Ray, 2018), a repository of surviving textual data from the pre-trial hearings of the *Salem Witch Trials*, to leverage new insights into the speech of American English in the late seventeenth century, contributing to an area of scholarship that remains relatively understudied (Kytö & Siebers, 2022). Trial data, including depositions, various letters and petitions, and recorded examinations, are used as diachronic surrogates of orality to reconstruct early American English in colonial Massachusetts. Using a historical data source like the *Salem Witchcraft Papers* necessarily cannot provide an accurate representation of 17<sup>th</sup> century speech (due to, for example, scribal interference), but may allow the researcher to reconstruct orality by choosing texts that most closely represent the spoken language (Kytö & Walker, 2003).

The application of the variationist framework (Labov, 1972; Weinreich, Labov, & Herzog, 1968) to the variable systems in question provides the opportunity to assess the contribution of social and linguistic factors to the selection of variants, elucidating the nature of the underlying grammar through the use of distributional and multivariate analyses. The thesis critically assesses previous scholarship on relativisation and *do*-support, identifying a number of methodological infelicities, including non-adherence to the Principle of Accountability (Labov, 1972), inadequate data (too few tokens, drawn from fiction or prose), and consideration of a limited number of possible contexts that *do*-support could be used in (e.g. affirmative declaratives, or negatives).

Among the key findings to emerge from the investigation of relativisation are: the statistically significant effect of complexity constraints on the selection of variants, disfavours the zero relative marker in more cognitively complex environments; the use of *which* as the only *wh*-relative to be employed to any appreciable extent; and the apparent favourable association of *wh*-relativisers with younger speakers. The use of relative marker *who* in the data is negligible, in contrast with findings based on contemporary mainstream varieties of North American English. For *do*-support, results suggest that the decline of affirmative declarative *do* at the end of the 17<sup>th</sup> century has been overstated in previous research (Ellegård, 1953; Kroch, 1989). Sentence type is the single most influential factor in the use of *do*-support, and interacts with every other factor group tested. *Do*-support is also subject to social conditioning in the 17<sup>th</sup> century, particularly age and gender.

This study has implications for the wider field of historical sociolinguistics by demonstrating that a comprehensive and empirically accountable analysis of relativisation and *do*-support can offer new insights into the social and linguistic conditioning of these variables in Earlier American English. Considered in the aggregate, the major findings of the thesis indicate that judicious exploitation of the *Salem Witchcraft Papers* confirms them as a unique resource for undertaking large-scale quantitative analysis of 17<sup>th</sup> century English and reconstructing antecedent varieties of English in their socio-historical context.

## ACKNOWLEDGEMENTS

It is difficult to know where exactly to begin, acknowledging the end of an academic journey that has spanned more than a decade.

Firstly, the people without whom this could never have happened. Chief among these is my supervisor, Stephen Levey. Stephen's unflinching persistence, precision, and endless patience for me were a guiding light through even the toughest moments of the thesis process. From the moment I expressed interest in applying for a PhD at the University of Ottawa, Stephen has been a constant and reassuring presence, and I can say without hyperbole that this would not have happened without him. I could fill a thousand pages with gratitude, and it still would not be enough.

Also fundamental to my journey at the University of Ottawa are Shana Poplack and Dennis Ott, who both shaped my academic experience and gave me the confidence to continue striving, pushing me out of my comfort zone and beyond.

I would also like to thank all of the professors at the University of Chester and University of Sheffield, who encouraged me to grow and follow my interests and passion.

In my personal life, there are many people who have supported me. In no particular order, and with any egregious absences the fault of memory rather than malice: Niela, who has never stopped believing in me, and pushing me to greater and greater heights; Hamid joon, Mojgan joon, Nora joon, who have been here every step of the way, always supportive and only occasionally distracting; the Ottawa Persians, who always expressed interest in what I was doing; the Cottage Crew for providing a much-needed escape every year; Roya and Thomas, for always hosting us in Toronto, and talking me down from the ledge numerous times; Len and Fred for the good times and the merriment; and my family back in England for always supporting me, particularly my parents and my siblings.

For Maman Parvin, whose approval set me on this road, and for Niela, who kept me on it.

## CHAPTER 1: INTRODUCTION

### 1.1 The 17th century as a linguistic period

The Early Modern English Period (1500-1700) is a crucial stage in the development of the English language, bridging the gap between Middle English (1100-1500) and Late Modern English (1700-Present) (Nevalainen, 2006). Linguistic study of Early Modern English has not been as extensive as for earlier periods such as Old and Middle English (Görlach, 1991), but there have been notable contributions targeting this period in the fields of historical sociopragmatics (Archer, 2005; Culpeper & Archer, 2008; Fitzmaurice & Taavitsainen, 2007; Jucker, 1995; Jucker & Taavitsainen, 2010; Kryk-Kastovsky, 2000) and historical sociolinguistics (Ball, 1996; Ellegård, 1953; Huber, 2017; Nevalainen & Raumolin-Brunberg, 2017; Rissanen, 1985, 2003; Romaine, 1982; Söderlund, 2017).

The Early Modern period witnessed major social and political developments, including the English colonisation of the Americas, widespread radicalism and religious dissent (Hill, 1995), the English Civil War and subsequent Interregnum (Braddick, 2009), and a shift from a rural, agrarian economy to a mercantile economy, leading to increased importance of port towns (and the varieties of English spoken therein).

The Early Modern English Period also accommodated many major linguistic developments, including phonological changes like the Great Vowel Shift, which began in the 1400s during the Middle English Period, and stabilised in the south of England in the 17<sup>th</sup> century (Nevalainen, 2006: 7). Syntactic changes include the simplification of verb inflections (particularly the loss of tense-marking in the present tense), leading to a more fixed word order; the decline in the use of the second person pronouns *thou* (informal, singular) and *ye* (plural); and the disappearance of multiple negation from the emergent standard written variety (Los, 2015; Nevalainen, 2006: 111-113; Rissanen, 2003: 87). The two linguistic variables targeted in this dissertation, *do*-support and relativisation, were also implicated in variation and change during the Early Modern Period, particularly the steady decline of *do*-support in affirmative declaratives (Ellegård, 1953; Rissanen, 1985; Söderlund, 2017), and changes to the system of restrictive relativisation, including the putative emergence of an animacy constraint restricting *who* to human antecedents, and *which* and *that* to non-human antecedents (Ball, 1996; Romaine, 1982).

## 1.2 Early American English

As noted above, Early Modern English has received notably less sustained attention than other time periods in the history of the language, and this is particularly the case for Early American English. While synchronic studies on American English are common, there is a distinct lack of diachronic research on early American English, particularly with a focus on the Early Colonial Period (Kytö & Siebers, 2022: 2). In part, this may be due to the scarcity of surviving textual evidence that approximates the spoken language of that period, rendering the reconstruction of earlier varieties of spoken English, such as 17<sup>th</sup> century American English, difficult (Kytö & Siebers, 2022: 6).

Extensive research has been conducted with the aim of identifying the source varieties of early colonial American English (Kurath, 1986; Kytö & Siebers, 2022: 4). Thompson (1994: 14, 17) identifies early settlers as coming from affluent eastern counties in England, including Cambridgeshire, Essex, Lincolnshire, Norfolk, and Suffolk. However, as has been noted by Schneider (2004: 263), it is difficult to reliably trace colonial varieties to their input varieties. Any surviving evidence should therefore not be considered conclusive proof of a relationship between colonial and input varieties.

## 1.3 Data in Diachronic Research

The use of textual evidence is well-established in historical linguistics, but the increasing importance and availability of electronic data over the past 30 years has rendered large corpora of historical data more accessible and inspired new approaches to diachronic research (López-Couso, 2016: 127-129). Such approaches include the use of multi-genre corpora (e.g. the *Helsinki Corpus*; *ARCHER*) to analyse the effects of register, and the production of corpora that are time-period or genre-specific (e.g. the *Corpus of Nineteenth Century English*; the *Corpus of Early English Correspondence* and its derivatives) (López-Couso, 2016: 133-134).

Traditional approaches to diachronic research have tended to focus on literary texts (e.g. Ellegård, 1953), but the increasing availability of electronic corpora have galvanised attempts to reconstruct the spoken varieties of the past through the use of speech-like genres as a proxy for actual speech (e.g. the *Corpus of English Dialogues 1560-1760*, the *Old Bailey Corpus*, and the *Salem Witchcraft Papers*). This includes

the use of speech-like genres such as trial data and witness depositions (Rissanen, 1985, 2003; Söderlund, 2017), and epistolary documents (Nurmi, 2018). The research reported in this dissertation aligns with this trend.

#### **1.4 Putting the *Salem Witchcraft Papers* into historical context**

In 1692, in the burgeoning Massachusetts colony, accusations of physical and mental harm caused by witchcraft led to the deaths of 20 people, and the imprisonment of more than 200 (Baker, 2015; Hoffer, 1997). Taking place between February 1692 and May 1693, the Salem Witch Trials were an important moment in Early American history that continues to solicit public interest. The original accusers were four young girls from Salem Village, and the wave of accusations eventually spread to nearby towns like Andover and Topsfield (Demos, 1983). Most of the accused were women, although some men (e.g. Giles Corey, George Burroughs) were also indicted (Hoffer, 1997).

The data under investigation are drawn from the digital edition of the *Salem Witchcraft Papers* (Ray, 2018), which contains records from pre-trial hearings, examinations, testimonies and depositions, letters and petitions, as well as official documents including indictments and warrants for arrest (Grund, 2007: 122). The *Salem Witchcraft Papers* are situated at the end of the Early Modern English Period (1692-1693) as it transitions into the Late Modern Period, distinguished by the stabilisation of many of the linguistic changes that began in Middle English, and an emergent written standard (Rissanen, 2003: 87). Such trial documents, which are the nearest available analogue of 17<sup>th</sup> century speech, are a rarity, and the *Salem Witchcraft Papers* are an as yet under-utilised resource for historical sociolinguistic research (Kytö & Siebers, 2022: 6; Kytö & Walker, 2003).

Much of the previous research targeting the *Salem Witchcraft Papers* has approached the data through the lens of historical pragmatics (Archer, 2002; Doty, 2007; Doty & Hiltunen, 2002; Grund, 2021; Hiltunen, 1996; Hiltunen & Peikola, 2007), though sociolinguistic work on the dataset does exist (Ball, 1996; Díaz-Vera, 1999; Kytö, 1991, 2004, 2022; Rissanen, 1997, 2003).

Among variationist approaches, Rissanen (2003) evaluates the *Salem Witchcraft Papers* as a fertile resource for sociolinguistic investigation, citing preliminary results relating to several linguistic variables

(*do*-support, *no/not any*, *shall/will*) in an early subset of the *Salem Witchcraft Papers*. Rissanen (2003: 111-112) concludes that the *Salem Witchcraft Papers* allow the researcher access to three key extralinguistic factors that are crucial to the variationist framework: sociolinguistic variability, or access to individuals from different social backgrounds; textual variability; and regional variability. Similarly, Schneider (2013: 90) lists several characteristics that render a historical dataset appropriate for variationist study. These include proximity to the vernacular spoken language through the use of speech-like text types (i.e. looking beyond formal and canonical literary writing); extralinguistic variation (different authors of varying age, sex, and social status, and stylistic variation); the presence and variability in the use of the target phenomena; and sufficient availability of data to allow for quantitative analysis, with large numbers of each variable (Schneider, 2013: 90). The *Salem Witchcraft Papers* meet these requirements.

One of the most prevalent issues in any study of historical texts is the ever-pervasive “bad data” problem (Labov, 1994: 11). The further separated one is from a historical speech act, the more limited the availability of data and associated socio-biographical information, and the greater the risk of scribal or editorial intervention. Unfortunately, any historical linguistic study is restricted to the data that has survived, and such survival is entirely random (Labov, 1972: 100; Nevalainen, 1999: 499). While scribal interference is unavoidable (and errors have been introduced in the transcription of the documents), the *Salem Witchcraft Papers* do excel in the preservation of sociobiographical data. The legal context that the data was recorded in has ensured that a decent amount of information about the individuals involved has survived.

In the study of both relativisation and *do*-support, access to speech-like data is crucial. One of the major claims about relativisation is that it has not made significant inroads into the spoken language (Romaine, 1982), and *do*-support has been claimed to originate in the spoken language (Rissanen, 1985). Without access to authentic historical speech, it is impossible to establish the fidelity of the dataset to the spoken language of the target time period. The nearest available analogue is through the use of written texts that purport to record spoken dialogue, such as trial examinations (Auer et al., 2015: 7; Kytö & Walker, 2003: 221; Schneider, 2013). As with all historical documents, caution should be exercised in directly associating the data in the *Salem Witchcraft Papers* with vernacular speech. Grund (2007: 121) observes,

based on scribal notation, that the focus in this dataset was on substance rather than literal context. Scribes may have recorded examination events in shorthand,<sup>1</sup> to be written in full at a later time (Grund, 2007).

As with all historical data, there is the potential for scribal interference (discussed in more detail in Section 2.3). The original recorder may not record the speech act verbatim, perhaps correcting or omitting non-standard usage (Kytö & Walker, 2003: 230). Even so, the limited nature of surviving historical data means that trial documents are still one of the most reliable resources for accessing historical speech-like data. It should be made clear that while the *Salem Witchcraft Papers* do offer many advantages to the researcher, they are not a faithful rendition of 17<sup>th</sup> century speech, and should be considered at most to be an approximation of spoken Early Modern English.

A colonial source like the *Salem Witchcraft Papers* offers a unique window into the language of 17<sup>th</sup> century Massachusetts (Kytö & Siebers, 2022: 2). It allows for comparison with contemporaneous British English data, highlighting early divergence between the two varieties, and can contribute to discussions around ‘colonial lag’, according to which transplanted varieties are more linguistically conservative than their source varieties (Hundt, 2009; Marckwardt, 1958).

## 1.5 Linguistic Variables Under Investigation

The linguistic variables that form the basis for the current research (restrictive relativisation and *do*-support) were selected because they have been reported to be undergoing change at the end of the 17<sup>th</sup> century (Ball, 1996; Ellegård, 1953; Rissanen, 1985, 2003; Romaine, 1982). Both variables have received attention in the literature, but heavy reliance on canonical literary data, coupled with the eschewal of key working principles of variationist sociolinguistics (e.g. the Principle of Accountability; Labov, 1972), are among the methodological infelicities in previous treatments that this dissertation seeks to highlight and address.

### 1.5.1 Relativisation

The Modern English system of restrictive relativisation, in which there is variation between the *wh*-relatives (*who*, *which*, *whose*), *that*, and the zero relative, has its roots in Early Modern English

---

<sup>1</sup> Though as Grund (2007: 31) notes, examples of shorthand versions of the examinations do not appear to have survived, and questions remain as to the quality and accuracy of the recording systems used.



(Romaine, 1980: 222). *That* and the zero relative have their origins in Old English (though the latter is reportedly rare), while the *wh*-relatives are widely acknowledged to be innovations of the Middle English Period (Fischer et al., 2000: 92; Romaine, 1982). The Early Modern Period is reported to be characterised by the development of an animacy constraint restricting *who* to human antecedents and *which* to non-human antecedents (Ball, 1996: 246), and the claimed ascension to dominance in written varieties of the *wh*-relatives (Fowler & Fowler, 1931: 80; Johansson, 2012, 2017; Romaine, 1982; Tottie & Johansson, 2015). Romaine (1982: 212) claims that while the *wh*-relatives have made significant inroads into the written language, they are still infrequent in the spoken language (see also Tagliamonte, 2013: 104; Tottie, 1997). However, this notion is challenged by Ball (1996: 242-243), who observes that Romaine's (1982) claims are based on dialects (Scots and Appalachian English) where *wh*-relatives are infrequent. Of particular relevance to this dissertation is the fact that Ball's (1996) observations are based in part on the quantitative analysis of variable relativisation strategies in a subset of the *Salem Witchcraft Papers*.

The end of the 17<sup>th</sup> century is well-situated to test the claims summarised above, including the animacy constraint, and the use of relativisation in texts that more closely approximate the spoken language (e.g. examinations), in comparison with less speech-like genres (e.g. official documents).

### 1.5.2 *Do*-Support

*Do*-support is a linguistic phenomenon that has attracted extensive scholarly attention in the historical (socio-)linguistic literature (Bækken, 2002; Ellegård, 1953; Moessner, 2017: 179; Rissanen, 1985, 2003; Söderlund, 2017). The first attested use of *do*-support is in Middle English, but it is really during the Early Modern English period that it is said to become a productive feature of the language. During this period, *do*-support has been reported to decline in affirmative declaratives, and steadily increase in negatives and interrogatives (Ellegård, 1953; Kroch, 1989; Ogura, 1993).

The end of the 17<sup>th</sup> century is a particularly interesting period for the use of *do*-support. For example, in affirmative declaratives *do*-support has been claimed to have effectively disappeared from the language by 1700 (Ellegård, 1953; Kroch, 1989). Examining the use of *do*-support in Early American English specifically is pertinent as it has been claimed by Rissanen (2003: 108) that in the 17<sup>th</sup> century affirmative declarative *do* was more frequent in American English than in equivalent British English data.

Despite the prodigious amount of work on this variable, methodological issues abound. Many previous studies restrict their scope to the use of *do*-support in particular syntactic contexts, including affirmative declaratives (Bækken, 2002; Ihalainen, 1991; Nurmi 1999, 2018; Rissanen, 1985; Wischer, 2008), and negatives (Díaz-Vera, 1999; Pérez, 1997; Warner, 2005). In their examination of habitual *do*-support in contemporary Samaná and Somerset English, Jones and Tagliamonte (2004) only consider past tense contexts. Additionally, in several instances (e.g. Nurmi, 1999, 2018) the rate of *do*-support is calculated through the use of normalised frequencies per 10,000 words, which fails to contextualise the use of *do*-support within its larger variable system (i.e. where it systematically alternates with contexts where *do*-support is not used). Similarly, Ellegård's (1953) influential data uses a sample from each text to estimate the rates of *do*-support and the simple form. Such approaches are problematic for the study of *do*-support as it has been claimed (e.g. by Rissanen, 1985) to be subject to a clustering effect, with increased density in emotive or intense passages. By extrapolating rates of *do*-support from a sample of each text, Ellegård (1953) risks under-reporting rates, as the clustering effect means *do*-support is likely not evenly distributed throughout each text.

## 1.6 The Variationist Approach

The approach and analysis reported here draw on the variationist framework, which works on the assumption that language is inherently variable (Labov, 1966, 1969, 1972, 1982; Weinreich et al., 1968). The systematic variation between semantically equivalent variants, and the social and linguistic conditioning of linguistic variables (derived through quantitative analysis), lie at the heart of the methodology (Hazen, 2007; Labov, 1963; Tagliamonte, 2013: 730). Particular attention is given to language as it is used in everyday speech (Milroy, 1992; Sankoff 1988).

The quantitative analysis of linguistic variables draws on several pieces of evidence – the statistical significance of factors influencing the selection of variants; the relative contribution of different factors to the selection of variants, as gauged by the distribution of variants across factors; and comparison with claims in the literature (Hazen, 2007: 70-71; Jones & Tagliamonte, 2004: 110; Tagliamonte, 2011). Of most importance to the variationist approach is the constraint hierarchy, or the ordering (from strongest to weakest) of each independent factor's influence on the use of the target variants. The ranking of constraints

is indicative of the underlying grammar, and key to the analysis of variation (Poplack & Tagliamonte, 2001).

One of the basic tenets of the variationist paradigm is the Principle of Accountability (Labov, 1972). At its core, this means that the researcher must collect all instances where the target variable could have been used, even if it was not. Non-adherence to this basic principle prevents the researcher from establishing comprehensively how the variants belonging to a variable system are used. The Principle of Accountability allows the researcher to contextualise the use of a specific variant as a proportion of all valid instances within the target dataset, relating it to the other variants with which it systematically alternates (Wolfram, 1991), demonstrating the propensity for a speaker to use that particular variant (Poplack & Tagliamonte, 1989). The Principle of Accountability also serves to reduce the chance of data being ‘cherry-picked’ (Hazen, 2007: 76).

The variationist paradigm was initially developed to target variation and change in contemporary varieties (Labov, 1963; Weinreich et al., 1968). Later developments applied the methodology to socio-historical research (Pintzuk, 2017; Romaine, 1982, 2016). Diachronic research is subject to additional constraints and considerations, as any relationship between historical varieties and their social contexts must be reconstructed by the researcher (Romaine, 2016: 21). The social significance of linguistic variables is not stable, and is liable to vary over time.

Access to different social strata is restricted by the social norms of the target period. Prior to widespread literacy, extant data is weighted in favour of males who are higher in the social hierarchy, and much of it was written by scribes (Romaine, 2016: 23). The social hierarchy was also subject to change throughout history, making it difficult for the researcher to reliably establish a given individual’s position in the social order. The social hierarchy in historical contexts must be reconstructed by the researcher based on the available information, which may be inconsistent or fragmented (Romaine, 2016: 24).

Some examples of the application of the variationist framework to socio-historical research include Kytö (1991), who compares the use of modal auxiliaries in Early Modern British and American English; the studies on Early Modern English (1680-1800) in Nevalainen et al. (2018); Bergs’ (2005) work on the

Middle English Paston Letters (1421-1503); and research on historical varieties of French by Poplack and St-Amand (2007) and Ayres-Bennett (2004).

Variationist approaches to relativisation include Romaine (1982), whose seminal work examines relativisation in Middle Scots; Ball (1996) who finds evidence for an animacy constraint restricting *which* to non-human antecedents and *who* to human antecedents; Huber (2017) who looks at relativisation in the *Old Bailey Corpus* (1720-1913); and Levey and DeRooy (2021), who report on the use of relativisation in 19<sup>th</sup> century American English.

Historical use of *do*-support has elicited less attention from variationists, but notable examples include Wischer's (2008) work on Early Modern affirmative declaratives; Warner (2005) and Díaz-Vera (1999) on negatives; Rissanen (1985, 2003), who focuses on *do*-support in Early American English; and Söderlund (2017), who looks at *do*-support in Early Modern English depositions.

## 1.7 Major Research Objectives

The research presented here targets several major research objectives.

- I. to reconstruct the variable systems of the target variables (relativisation and *do*-support) in Massachusetts at the end of the 17<sup>th</sup> century, to investigate broader claims in the literature relating to the status of those variable systems in Early American English.
- II. to confirm, refute, and refine (e.g., by targeting methodological infelicities in previous treatments) extant claims in the literature regarding the social and linguistic conditioning of relativisation and *do*-support. Major claims about relativisation include Romaine's (1982) assertion that the *wh*-relatives have made limited inroads into the spoken language, and the claim advanced by Ball (1996) that during the Early Modern period relative markers became subject to an animacy constraint restricting *who* to human antecedents and *which* to non-human antecedents. For *do*-support, particular attention is given to claims regarding the virtual disappearance of affirmative declarative *do* by the end of the 17<sup>th</sup> century (Ellegård, 1953; Kroch, 1989), and lexical effects, where certain verb collocations have been reported to resist the use of *do*-support (Ellegård, 1953; Engblom, 1938; Nurmi, 2018; Warner, 2017).

- III. to evaluate the use of the *Salem Witchcraft Papers* as a linguistic resource for reconstructing antecedent varieties of English, building on the growing literature on the use of historical data for sociolinguistic research (Kytö & Siebers, 2022; López-Couso, 2016; Rissanen, 2003).

As the above sections have explained, the target variables, methodological framework, and dataset were deliberately and carefully selected to best address these research objectives.

The remainder of the thesis will be structured as follows. Chapter 2 will present a critical review of the *Salem Witchcraft Papers*, situating the data in their historical context, and justifying the application of the variationist methodology. Chapters 3 and 4 present the results of the investigations into the use of relativisation and *do*-support respectively, situated in their wider sociolinguistic context. The results include distributional analysis of numerous social and linguistic factor groups, and statistical verification of findings. Chapter 5 concludes the discussion by contextualising the quantitative patterns observed in the data with regard to both variables.

## CHAPTER 2: CRITICAL EVALUATION OF THE DATA AND THEORETICAL FRAMEWORK

### 2.1 The Salem Witch Trials

The *Salem Witch Trials* were a pivotal moment in American history, and to a lesser extent, European history. They signified the end of an era – the era of witch hunts, which had raged in Europe for nearly three hundred years, resulting in the deaths of thousands (Baker, 2015: 7). This particular conflagration began in the eponymous village of Salem, in the state of Massachusetts, but soon grew to encompass the surrounding area, and nearby towns in the same state, such as Andover and Topsfield, were also drawn into the hysteria. The first accusations were made in February 1692, and the trials continued until May 1693, when Governor William Phipps intervened, halting any further arrests and disallowing the future use of spectral evidence. The notion of spectral evidence was based on the assumption that someone in league with the Devil could give permission for their spectral form to be used to recruit or torment the innocent. The (controversial) use of spectral evidence is a defining feature of the *Salem Witch Trials* (Grund et al., 2004: 146; Rissanen, 2003: 85).

More than 200 people were accused of witchcraft, and 20 people were executed (Rissanen, 2003: 85). Several people perished while languishing in prison, and even those who survived relatively unscathed faced economic hardship as a result of the accusations. In the aftermath of the trials, some of the charged had their names cleared, and money was given to their families (Blumberg, 2007; Baker, 2015). The trials left an indelible mark on the culture of the region and continue to garner interest, as shown by the popularity of fictional representations like Arthur Miller's *The Crucible*.

Various explanations have been offered for the events that transpired during this period, including toxic fungi that caused hallucinations (Woolf, 2000), psychological hysteria after clashes with the indigenous population (Blumberg, 2007), and even attention-seeking on the part of the accusers (Kocić, 2010). The most compelling explanation, however, is that these events were rooted in a conflict between two key families in the region, the Putnams and the Porters, which spurred the initial wave of accusations. Many of those who were accused belonged to or had some connection to the Porters, and many of the accusers were connected to the Putnams. Tensions between the two families had been present for some time, but it wasn't until accusations of witchcraft began to proliferate that events turned dangerous. The tension

between these two families is connected to wider tensions between the primarily agrarian and religiously minded Puritans, and the more secular mercantile class that was becoming increasingly common in the area at the time. Grund (2007) cautions, however, that the witch trials are not reducible to a single underlying cause, and that they are likely the result of a combination of factors.

The legal process in the Massachusetts colony in 1692 was based on the Massachusetts Code of 1648, a codified set of laws written specifically for the colony, and the legal model for later colonies (Haskins, 1954: 3-4). While based on English common law at the time, the Massachusetts Code was distinct and specifically tailored to the legal needs of a colonial state. Typically, cases would be initiated by the victim providing the local magistrate with a formal accusation. In Salem, this was followed by preliminary examinations, which the extant examination data is drawn from (Doty & Hiltunen, 2009: 459; Grund et al., 2004: 150). These preliminary hearings served the function of determining whether the case was eligible to go to trial. If the case was determined to have enough evidence for trial, depositions from witnesses (both for and against) were collected. In Salem, these depositions were typically written outside of court, then submitted to the authorities (Grund, 2012: 6). As a result, there was no requirement that depositions be written by an official, and many were written by the deponent themselves. Indictments are highly conventionalized documents that officially charge the suspect with criminal acts, and were issued after depositions were collected (Doty & Hiltunen, 2009: 460). The initial accusation would also have been recorded as a deposition prior to indictment. Finally, the suspect underwent trial in a criminal court, which in Salem was known as the Court of Oyer and Terminer. Examination records from the Court of Oyer and Terminer have unfortunately not survived (Doty & Hiltunen, 2009: 460). Those who maintained their innocence, but were found guilty by the court, were often executed (Doty & Hiltunen, 2009: 466). Ironically, if the accused confessed to their crime and showed repentance, they were more likely to survive.

The court admitted three primary types of evidence in their search for proponents of witchcraft. First, and most infamously, was the admission of spectral evidence, which even at that time was controversial. However, as Craker (1997: 332) notes, the role of spectral evidence in the successful conviction of individuals accused of witchcraft has possibly been exaggerated. The second possible type of evidence is what Craker (1997: 332) terms “*non-spectral acts of malefic witchcraft*”, which encompasses everything from the ability to cast spells, curse people, and cause sickness and death, to telling the future and using

poppets and potions. Poppets, or puppets as they were sometimes called, were small dolls made of rags, often with pins stuck in them (Ray, 2003). They were at the time considered to be a tool used by witches to afflict people, akin to the “voodoo doll” depicted in Western media. Finally, confessions by the accused were admissible as evidence, though claims have been made that confessing (and not retracting the confession) was a guaranteed way to avoid execution (Craker, 1997: 33).

## 2.2 The *Salem Witchcraft Papers*

The *Salem Witch Trials* took place during a highly important period in the history of the English language. Towards the end of the 17<sup>th</sup> century, when the *Salem Witch Trials* took place, the language had undergone a period of rapid change, culminating in increased standardisation, and a reduction in the structural and morphological complexity and variability of written Middle English (Rissanen, 2003: 86-87).

Extant textual data from the period in question has been collated in Boyer and Nissenbaum’s *The Salem Witchcraft Papers: Transcripts of the Legal Documents of the Salem Witchcraft Outbreak of 1692* (henceforth *Salem Witchcraft Papers*), originally published in 1977. Almost all of the surviving records of the period are from pre-trial hearings, examinations and depositions (Grund, 2007: 122). The extant records provide a valuable insight into the language of early American speakers, and their importance as a linguistic resource cannot be overstated (Grund, 2007: 120; Rissanen, 2003: 86-87). The records allow modern researchers a window into the lives of the varied people in 17<sup>th</sup> century Massachusetts, and most relevantly, into the language that they spoke.

The data which form the cornerstone of this dissertation were drawn exclusively from the digital edition of the *Salem Witchcraft Papers* (Ray, 2018).<sup>2</sup> The digital edition differs from the Boyer and Nissenbaum (1977) edition in the inclusion of newly-located documents absent from the original edition. The digital edition of this corpus was assembled under the guidance of Professor Benjamin Ray at the University of Virginia. Transcription of documents was supervised by Professor Bernard Rosenthal, University of Binghamton. The digital edition became available in 2011, and was intended as a re-transcription of Boyer and Nissenbaum’s (1977) work, accounting for errors and omissions. Some of the amended errors include

---

<sup>2</sup> <https://saalem.lib.virginia.edu/category/swp.html>



incorrect dates, ages and names, inclusion of words that were missed in the original transcriptions, and correcting some words that were mistranscribed, such as ‘rats’ for ‘cats’. However, the *Salem Witchcraft Papers* were not transcribed with linguistic research in mind, and numerous transcription errors are still present in the data. This aptly demonstrates the need for caution when approaching historical data sources.

When dealing with a digitised version of historical data, it is important to identify and highlight to what extent the texts have been edited. With the digital edition of the *Salem Witchcraft Papers*, minor changes have been made. For example, superscript letters have been aligned with the rest of the text, and apostrophes inserted to indicate contraction. Example (1) gives a sentence as it appears in the original manuscript, while example (2) shows the sentence as it appears in the digital version of the *Salem Witchcraft Papers*. It is also worth noting one of the transcription errors repeated in example (2) – the first instance of *he* is erroneously omitted in the *Salem Witchcraft Papers*.

(1) he told me he put it in a purs among Som other mony w<sup>ch</sup> he locked vp in a box

(2) he told me put it in a purs among Som other mony w'**ch** he locked up in a box

(SWP013.10 – Samuel Shattuck,  
Deposition: 2<sup>nd</sup> June 1692)

Other than minor orthographic changes, the transcriptions retain the irregular spelling and idiosyncratic written conventions of 17<sup>th</sup> century written English. For example, *that* is often written (and transcribed) as *yt*.

(3) Liu<sup>t</sup> Rich<sup>d</sup> Hunniwell & John Greinslett were then present & some others y<sup>t</sup> are dead.

(SWP022.33 – Thomas Greenslit,  
Deposition: 15<sup>th</sup> September 1692)

A small number of tokens were erroneously excluded on the basis of transcription errors in the *Salem Witchcraft Papers*. While not enough to impact distributional percentages, this may result in frequency mismatches between this thesis and the manuscript. The accuracy of all retained tokens has been confirmed through cross-reference with the *Records of the Salem Witch-Hunt* (Rosenthal & Adams, 2009),

which aims for a more faithful reproduction of the documents, correcting transcription errors present in the *Salem Witchcraft Papers*. The totals reported in the results sections in Chapters 3 and 4 should therefore be considered minimum values, accounting for erroneous exclusions and human error.

There are nearly 1000 surviving records in the *Salem Witchcraft Papers*. Grund, Kytö and Rissanen (2004) calculate that around half (47%) of these are depositions. Petitions and examinations account for 8% of records each (though it should be noted that the examinations are typically greater in length than other text types). The remainder of records are of the highly formulaic type – warrants, indictments, recognizances,<sup>3</sup> and more (Rissanen, 2003: 88). Depositions tend to be less formulaic than warrants and indictments, but more formal than examinations, which record direct and indirect dialogue (Rissanen, 2003: 88). Depositions were typically recorded outside of the court, then submitted for inclusion as evidence in the trial. They could be written by anyone – including the deponent themselves, or their family and friends (Grund, Burns, & Peikola, 2014: 44-45).

The trials are somewhat unique in terms of the wealth of textual data that they left behind, representing speech-like data from a diverse range of people in terms of age, gender, and class (Grund et al., 2004: 146). For example, while many of the speakers belong to the mercantile class or are labourers, there is also data from slaves (Tituba, Candy, and John Indian), highly educated officials like William Phipps, and members of the clergy, like the Reverend Samuel Parris.

### 2.3 Reconstructing Orality

A major methodological consideration when working with historical data (particularly, but not exclusively, when working in a variationist framework), is how well the available data can be used to reconstruct *spoken* language as it was used in the target time period. The earliest known device capable of recording speech was Édouard-Léon Scott de Martinville's phonautograph in the 1850s (Brock-Nannestad & Fontaine, 2008). Prior to this invention, the sole trace of the spoken language lies in written texts (Culpeper & Kytö, 2010: 2).

---

<sup>3</sup> A form of appearance bond similar to bail, a recognizance was a guarantee from the accused that if they did not appear in court, they would face a financial penalty (Samaha, 1981).

The reconstruction of the speech of the past through textual evidence merits careful consideration. It is highly unlikely that a written representation of speech will perfectly mirror actual speech. Instead, the original speech act is ‘filtered’ through the written record, and the recording context may obscure or distort it (Schneider, 2013: 58). The crucial challenge for historical sociolinguistics, then, is assessing and establishing in a reasonable and justified manner exactly how much distortion has taken place, and how well the written record reflects the original speech act (Montgomery, 1997: 227). The texts chosen must also feature the phenomenon of interest, and in such a way that there is variability in the targeted phenomenon (Schneider, 2013: 60).

Assessing the fidelity of the extant data to the original speech act is a particularly challenging issue. Without an authentic contemporaneous *spoken* baseline, establishing the fidelity of the historical speech-surrogate to the original speech act is an impossible task. Instead, a researcher must use the available material (i.e. speech-surrogate data) to reconstruct spoken varieties from the past. When data are being used to investigate the speech behaviour of a specific community, with only a written record surviving, it can be difficult to know if features in the data are reflecting vernacular speech, scribal idiosyncrasies, mistakes, or, indeed, some combination thereof. The scribal influence on the data likely plays at least some role in the accuracy of the representation of speech. For example, certain aspects of language, particularly phonology, are more difficult to represent accurately. Another possibility is that non-standard linguistic features are being obscured (i.e. “corrected” by the recorder) through the recording or editing process. Surviving textual data is therefore at best an *approximation* to actual speech, and certain legal genres (e.g. examinations) represent speech better than others.

When dealing with historical data, the closest one can get to authentic speech is through written texts that attempt to record spoken dialogue (Auer et al., 2015: 7; Kytö & Walker, 2003: 221). Records from trial proceedings, such as examinations and depositions, are excellent sources of this sort of data. Such data is believed to be an approximation of the original speech event, as in many cases the recording is completed at the time of the speech event (Schneider, 2002: 73). Although they are valid and useful records of historical language data, and in many cases the closest possible approximant of historical speech, the use of trial data necessitates caution on the part of the researcher. Written recordings of original speech acts (such as examinations), even if recorded at the time of the speech act, cannot be held to portray verbatim speech. While lexical and syntactic structures might be accurate, many features of spoken language are at

best very difficult to record in writing (Kytö & Walker, 2003: 221). These texts are necessarily processed through a scribal interface, with the possibility that the scribe “corrects” what is being recorded (Kytö & Walker, 2003). In some cases, the speech act was initially recorded in shorthand, adding an additional layer of editorialising (Boyer & Nissenbaum, 1977). In the *Salem Witchcraft Papers*, scribal interference can be seen through the presenting of passages of speech in the third person, as the example below demonstrates.

(4) Susan: Sheldon & Mary Warren testified that now his appearance comes from his body & afflicts them.

(SWP137.5 – Samuel Parris,  
Examination: May 18<sup>th</sup>, 1692)

The possibility of scribal interference, whether conscious or not, must be considered in any sociolinguistic exploration of historical data. Particularly when it is indirect speech that is reported, the distance between the speech event and the written record is increased. Each layer of editing and copying that recorded speech goes through removes it further from the original speech event, and introduces the possibility for errors that do not accurately represent the spoken language of the time period (Kytö & Walker, 2003: 230). Research on spoken language as represented in written records should therefore not equate these data with actual vernacular speech, but must attempt to identify texts that are likely to be as faithful as possible to the actual speech event.

The use of trial data to reconstruct orality is well established in the broader field of historical linguistics, and related sub-fields (Auer et al., 2015: 7). For example, much work has been done in historical (socio-)pragmatics (Archer, 2005; Culpeper & Archer, 2008; Kryk-Kastovsky, 2000). In historical sociolinguistics, researchers like Ball (1996) and Huber (2017) have used trial data to explore variability in relativisation strategies. The *Salem Witchcraft Papers* themselves have attracted attention as a valuable linguistic resource in efforts to reconstruct orality (Archer, 2002; Doty, 2007; Doty & Hiltunen, 2009; Grund, 2007; Rissanen, 2003). However, this resource has been under-utilised in the reconstruction of historical morpho-syntax, affording an excellent avenue for variationist research.

In the case of the *Salem Witchcraft Papers*, scribal notation indicates that the focus was on substance over literal context (Grund, 2007: 121). The records of examinations, for example, appear to be reconstructions of speech events based on notes or shorthand, and may best be considered a blending of the language of both speaker and scribe. Where speech is recorded as a dialogue between speakers, the inclusion of speech-related features (discourse markers, interjections etc.) may indicate closer proximity to the original speech event (Grund, 2007: 122). The recorded speech may not reflect the spoken language of the time so much as it reflects the recorder's interpretation of spoken language at the time.

While trial proceedings are often cited as an ideal resource for historical speech, their status as semi-formal texts merits consideration. The original speech acts primarily took place in formal settings (e.g. in court), and so the written representations contain varying amounts of formulaic data (Culpeper & Kytö, 2000: 195; Doty & Hiltunen, 2009: 459; Grund, 2007: 210; Grund, 2012: 8-9). These formulae are an integral part of the legal genre, and serve to organise the information in legal cases along predictable pathways. However, as noted by Van Herk and Poplack (2003: 242) in their examination of letters written by antebellum African American settlers in Liberia, the presence of formulaic language (in the form of epistolary salutations) does not necessarily preclude the use of non-standard constructions.

The various genres in the *Salem Witchcraft Papers* are subject to different levels of formulaic language, and can be placed rather neatly along a continuum. Doty and Hiltunen (2009) discuss three text types - indictments, depositions, and examinations. While these do not encompass the full range of text types in the data, they do represent three broad categories - fully formulaic, partially formulaic, and non-formulaic. The text types not mentioned by Doty and Hiltunen (2009; e.g. recognizances, letters, petitions, etc.) can be reasonably placed under one of the above categories.

At the most formulaic end of the spectrum are documents that represent formal written legal discourse. Doty and Hiltunen (2009) discuss indictments as an example of this in the Salem data - other text types include recognizances, summons, warrants, and mittimus.<sup>4</sup> These texts typically follow very rigid

---

<sup>4</sup> A mittimus is a specialised legal document issued to a sheriff to command them to arrest an accused person and deliver them to jail (Burns & Rosenthal, 2008: 406).

structures, with formulaic wording that shows very little variability, and none of the variation associated with the spoken language (Doty & Hiltunen, 2009: 468).

Examinations occupy the least (or non-) formulaic end of the spectrum. There is little in the way of formulaic language, although the questions asked by the examiner may be similar in nature between examinations. These texts can be assumed to most closely represent the spoken language of the time (though they are still subject to scribal and editorial interference). There are certain structural patterns that examinations typically adhere to, which Doty & Hiltunen (2009: 465) outline as follows:

1. Accusation > denial
2. Accusation > confession
3. Accusation > denial > confession

As these structures apply at the textual level (rather than clausal or sentential), it is unlikely that there will be much influence on the presence of variation at the syntactic level (Doty & Hiltunen, 2009: 466).

Finally, depositions sit in a hybrid territory, and could be considered semi-formulaic. The opening and closing of a deposition, which introduce the deponent, the authority of the deposition, and its context, follow formulaic clause patterns.

The remainder of the deposition, termed a narrative by Doty and Hiltunen (2009: 462), represents more closely a written version of the deponents' actual words. The narrative section is typically (though not always) written in the first person, and uses informal, non-formulaic language, with little in the way of legal phraseology. The majority of depositions in the *Salem Witchcraft Papers* contain speech-like narrative sections, and are therefore a valuable resource for investigating variation.

- (5)           he did tortor me most cruelly by beating me with the Two sticks which he had in his hands  
(SWP078.11 – Mercy Lewis,  
Deposition: 4<sup>th</sup> August, 1692)

(6) I was present with him w<sup>n</sup> he dyed & I do aferm that he dyed peacibly and quietly

(SWP016.15 – William Carr,

Testimony: 9<sup>th</sup> September, 1692)

## 2.4 Theoretical Framework

This research utilises the variationist approach (Labov, 1972; Weinreich et al., 1968) to investigate restrictive relativisation and *do*-support in the *Salem Witchcraft Papers*. This is a theoretical framework which takes the linguistic variable as its core analytical component. A linguistic variable is made up of competing variants which are more or less interchangeable insofar as they express the same referential meaning or a similar grammatical function. This interchangeability, or variation, is at the heart of the variationist approach, which is predicated on the assumption that the choice between two (or more) variants is far from random, and is in fact constrained by both internal (i.e. linguistic) and external (i.e. social) factors (Tagliamonte, 2013: 730). A key working tenet of this approach is that the ‘grammar’ underlying surface patterns of variation can be inferred from detailed inspection of the distribution and conditioning of competing variants in discourse (Poplack and Levey, 2010: 398). Quantitative analysis enables the objective establishment of patterns in the data, which can be used to inform claims about the conditioning of factors and to identify the constraints governing variant selection. This enables the researcher to systematically model the grammatical and social constraints governing linguistic variables.

In order to elucidate the contribution of independent variables to the use of relativisation and *do*-support, a combination of distributional and multivariate analysis is used. Mixed-effect logistic regression is carried out using Rbrul (Johnson, 2009). The statistical significance of an independent factor or predictor (at the 0.05 level) indicates whether the predictor exerts a significant effect on variant selection (Johnson, 2009; Poplack and Tagliamonte, 2001: 92; Tagliamonte, 2011: 122). Of utmost importance is the direction of effects, or constraint hierarchy, which shows the order of factors within a factor group. The ranking of constraints (or factors) reflects the grammar underlying the variation, and serves as the key diagnostic in assessing the nature of the underlying grammar. (Poplack and Tagliamonte, 2001: 94).

The current research supplements the variationist work that has been carried out using the *Salem Witchcraft Papers* (e.g. Ball, 1996; Kytö, 1991, 2004, 2022; Rissanen, 1985, 2003), and contributes to the

wider field of historical sociolinguistics. By exhaustively extracting valid tokens in accordance with the Principle of Accountability (Labov, 1972), and testing numerous linguistic and social factor groups abstracted from the literature, it is possible to gain some insight into the factors conditioning the use of the two target variables, restrictive relativisation and *do*-support, in 17<sup>th</sup> century English. Statistical testing strengthens the results, and allows corroboration and refinement of extant claims.



## CHAPTER 3: RELATIVISATION

### 3.1 Introduction to Relativisation

The variable of interest in this chapter is the relative system, and particularly restrictive relative clauses. One of the best studied syntactic variables (Ball, 1996; Bayley, 1999; Herrmann, 2005; Hinrichs et al., 2015; Huber, 2017; Johansson, 2017; Levey & Hill, 2013; Romaine, 1982; Tagliamonte et al., 2005; Tottie & Rey, 1997), the system of restrictive relativisation comprises a set of variants, with the primary alternation being between *that*, *wh*-relativisers (such as *which* and *who*), and the zero relativiser. In spite of the prodigious amount of attention that relativisation has received, previous studies have been hampered by a lack of empirical adequacy. For example, some earlier studies (e.g. Beaman, 1984; Ihalainen, 1980; Johansson, 2006; Schneider, 1989) likely over-estimate rates of *wh*-relativisation by not separating restrictive and non-restrictive relative clauses, which are structurally distinct. Non-restrictive relative constructions are reported to largely favour the use of *wh*-relatives. Consequently, any failure to systematically distinguish non-restrictive from restrictive relative clauses risks artificially inflating overall rates of *wh*-relativisation.

The primary objective of this chapter is to use diachronic surrogates of orality to determine the distribution and sociolinguistic conditioning of variants in the relative system as it was used in 17<sup>th</sup> century Massachusetts English. I also seek to address extant methodological issues in the literature, building on and refining previous findings. Conducting a thorough, accountable analysis of historical data using a variationist framework allows the possibility for direct comparisons with other research on the relativiser system (e.g. Ball, 1996; Huber, 2017; Johansson, 2017; Levey and DeRooy, 2021).

### 3.2 History of Relativisation in English

#### 3.2.1 Key Works

The relative system in English has received much attention in the literature from both synchronic (e.g. Hinrichs et al., 2015; Romaine, 1981; Tagliamonte et al., 2005) and diachronic perspectives (e.g. Ball, 1996; Huber, 2017; Johansson, 2017; Romaine, 1982). Research on relativisers has come from various sub-fields of linguistics, including syntax (Chomsky, 1956; De Vries, 2002) and first and second language acquisition (De Villiers et al., 1979; Diessel & Tomasello, 2005; Flynn et al., 2004; Guasti &

Cardinaletti, 2003). The relative system has also received attention in typological research (e.g., Keenan and Comrie, 1977). Following Romaine's (1982) seminal study of relativisation in Middle Scots, articulated from a diachronic sociolinguistic perspective, relativisation has garnered much traction in the sociolinguistic literature (Ball, 1996; Bayley, 1999; Herrmann, 2005; Huber, 2017; Tagliamonte, 2002; Tagliamonte et al., 2005; Tottie & Rey, 1997). There still remains a number of questions to be elucidated, and the English relative system continues to offer fruitful avenues for research. Of most relevance here is the opportunity for accountable, quantitative historical studies using diachronic surrogates of orality, situated in a variationist framework.

The general consensus in the literature is that in Present-Day English, the relative system is based on a tripartite distinction between the *wh*-forms (*which*, *who(m)*, *whose*), the much older form *that*, and the zero relative (Beal & Corrigan, 2007: 103; Tagliamonte, 2013: 94; Tottie & Rey, 1997: 219). Regional variants such as *what*, *as*, and *at* are reported to be considerably more marginal in Standard English, and do not really compete to any appreciable extent with the *wh*-relativisers, *that* and *zero* (Tagliamonte et al., 2005), though as Cheshire (1982) notes, *what* is quite productive in varieties like Reading English. Contemporary examples of each from Tagliamonte et al. (2005: 76) are given below. In what follows, the symbol  $\emptyset$  is used to designate the null or zero relativiser.

- (7)
- a. Beside me was the lady who presented the queen.
  - b. Building the houses which we know now as Church-Terrace.
  - c. They were good herring that we got.
  - d. But there were a boy in Ballyclare  $\emptyset$  told me this.
  - e. I would have carred the groceries what they needed.
  - f. Aye, well, I likes folk as was in it.

The relative system has experienced substantial changes in the history of the English language. The major changes in each period of English will be discussed in detail below, with particular attention given to the situation in Early Modern English, the focus of the present study. Broadly, the major changes can be summarised as follows: the domination of *that* in the latter stages of Old English; the gradual appearance of the *wh*-relatives in Middle English; the development of the animacy constraint in Early Modern English,

favouring *who* with human antecedents and *which* with non-human antecedents (Ball, 1996: 246; Johansson, 2012: 784-786), and the putative ascension to dominance of the *wh*-relatives (Fowler & Fowler, 1931: 80). The relative system in Present Day English is reported to largely resemble the relative system in Early Modern English (Romaine, 1980: 222), though the extent to which this is true awaits empirical corroboration, a point to which I return in the discussion of the findings.

Romaine's (1982) seminal study was an important milestone both in the development of research on the relative system, and in terms of the extension of the variationist paradigm to historical syntax. Much subsequent research on English relatives has been inspired or driven by questions that were raised in that study. In particular, Romaine's (1982: 212) claim that the *wh*-forms have successfully infiltrated the written language, but have hardly affected the spoken language has served as major impetus for additional research and debate. Romaine's (1982) original claim has been widely discussed in the sociolinguistic literature (e.g. D'Arcy & Tagliamonte, 2010; Geisler, 2002; Tagliamonte, 2002; Tagliamonte et al., 2005; Tottie, 1997). For example, Tagliamonte (2013: 104) found very low rates of *wh*-forms, particularly in vernacular dialects of British English.<sup>5</sup>

By contrast, in their study on restrictive relatives in spoken Tyneside English Beal and Corrigan (2007: 107) disagree with Romaine's claim, stating that the *wh*-relatives have infiltrated the spoken language, particularly with animate subject antecedents. Ball (1996: 243) also challenges Romaine's (1982) claim, citing regional speech data from Cheshire (1982; Reading), Macaulay (1991; Scots), and Hackenberg (1972; Appalachian English). Ball (1996: 242-243) asserts that the dialects of Scots and Appalachian English, which were the basis for Romaine's (1982) claims, are "essentially WH-less". Cheshire's (1982) data on the English of working-class adolescents in Reading, Berkshire shows quite high rates of *who* and *what* in personal and non-personal subject relatives respectively. For Ball (1996), this counters Romaine's (1982) assertion that the *wh*-forms have not infiltrated the spoken language.

---

<sup>5</sup> Tagliamonte et al. (2005) examined relativiser selection in several dialect areas: Northern Ireland (Portavogie and Cullybackey), Lowland Scotland (Cumnock), and Northwest England (Maryport). In Tagliamonte (2013), this data is analysed alongside comparable data from Northeast England (York, Wheatley Hill) and Southwestern England (Tiverton in Devon, and Wincanton in Somerset). Rates of *wh*-forms were low across all dialects, but were slightly higher for York and Wheatley Hill. These Northeastern English dialects are distinct from the others examined primarily in terms of their proximity to urban centres. Tagliamonte (2013) connects this proximity with higher rates of *wh*-forms.

This in turn leads to Ball's (1996: 243) claim that "there is no vernacular norm [...] with respect to the distribution of relative markers in personal and nonpersonal subject relatives". This is an issue that has been taken up in the literature as well. Bayley (1999: 119) connects the use of *wh*-forms with the standard language, and the dearth of *wh*-forms with vernacular varieties. D'Arcy and Tagliamonte (2010: 383) note that studies reporting higher rates of *wh*-forms in speech typically focus on standard varieties. Tagliamonte (2013: 103), looking at speech data in several varieties of British English, found higher rates of *wh*-forms in urban dialects. Her findings link the prominence of *wh*-forms not to any region, but to proximity to standard varieties of English. Non-urban varieties are found to resist the infiltration of *wh*-forms in restrictive relative clauses, confining them to non-restrictive relative contexts.

Britain (2021: 110-111), however, in an exploration of relativiser selection in several locales in East Anglia, finds that an urban hierarchy is not a suitable method of predicting adherence to standard linguistic norms. The distribution of relative markers in varieties of East Anglia English shows that more urban varieties do not necessarily pattern alike with regards to the distribution of relative markers. For example, Holbrook, a small village in East Anglia, shows high rates of *who* equivalent to those found in the urban city of Colchester. Britain (2021: 111) instead connects this pattern with a current trend of counterurbanisation, where (predominantly educated, middle-class) speakers are increasingly moving from urban centres to very rural areas.

The intersection of these two themes (the incursion of *wh*-forms into the spoken language, and the lack of a vernacular norm), forms a strategic foundation for the current study on relativisation in 17<sup>th</sup> century English.

This chapter targets restrictive relative clauses in 17<sup>th</sup> century American English, using the documents from the *Salem Witchcraft Papers* as a window into the spoken language of the time. Here, the value of the trials as a linguistic resource is of particular relevance. As a syntactic variable, relativisation is relatively rare in running discourse, meaning that a large amount of data is required to obtain enough tokens for effective quantitative analysis. Being official records, the trial data encompass a range of metadata about speakers, the dates of recording, and many aspects of the recording context (including who recorded witness depositions). Furthermore, the materials have been subjected to relatively little editorial intervention by the compilers of the corpus (Rosenthal & Adams, 2009).

In order to situate this research focus in its larger historical context, the remainder of this section will sketch the development of the relativiser system from Old English to Early Modern English. The major milestones in each period will be highlighted. Particular attention will be paid to the relative system in Early Modern English, as this is the target period of the present study.

### 3.2.2 Relativisation in Present Day English

As stated earlier, in Present Day English, the relative system is considered to be based on a tripartite distinction between the newer *wh*-forms, and the older forms, *that*, and the zero relative (examples from Guy & Bayley, 1995: 148).

- (8) a. This is the house *which* I told you about.
- b. This is the house *that* I told you about.
- c. This is the house  $\emptyset$ I told you about.

Relative clauses can be broadly divided into two categories – restrictive relative clauses and non-restrictive relative clauses. Examples of each from Tagliamonte et al. (2005: 76, 85) are given below.

#### *Restrictive relatives:*

- (9) a. Beside me was the lady who presented the queen.
- b. Building the houses which we know now as Church-Terrace
- c. They were good herring that we got.
- d. But there were a boy in Ballyclare  $\emptyset$ told me this.

#### *Non-restrictive relatives:*

- (10) a. We had a doctor last week, um basic doctor from Aspatria, which was very, very interesting.
- b. And John-Docherty, who died fishing, was there.

It is important to distinguish between the two types of relative clause for several reasons. Syntactically they are different (Traugott, 1992: 223). For example, restrictive relative clauses require nominal (either a pronoun or an NP) antecedents, while non-restrictive relative clauses have access to a wider range of possible antecedents, such as a VP or a PP. Examples of non-restrictive relative clauses with non-NP antecedents are given in Arnold (2007: 274), reproduced below.

- (11)       a. Kim [won the race], which I didn't think she could. (VP)  
              b. Kim put it [on his back], which was the right place. (PP)

Similarly, restrictive relative clauses and non-restrictive relative clauses are contrasted semantically (Tagliamonte et al., 2005: 85). Non-restrictive relative clauses introduce supplementary information, where restrictive relative clauses refer to an already mentioned antecedent. Examples from Tagliamonte et al. (2005: 86) given below.

- (12)       a. I went for mi <sub>(sic)</sub> pay, and I got twenty-five-bob, *which* is one-pound-twenty-five today.  
              b. She had three sisters, *who* were all married and away.

They also do not admit the same variation. Non-restrictive relative clauses overwhelmingly favour *wh*-forms, and do not ordinarily allow the zero relative (Guy & Bayley, 1995: 148; Suárez, 2012: 90; Tagliamonte, 2013: 96-97; Tagliamonte et al., 2005: 85). The relative marker *that* only rarely introduces non-restrictive relative clauses (Huber, 2017: 79-80). On the basis of the above information, it seems clear that restrictive and non-restrictive relative clauses are subject to different constraints, and represent different variable systems. They should therefore be examined separately. Results from studies that do not separate restrictive and non-restrictive relative clauses should at best be viewed cautiously, and will likely over-estimate rates of *wh*-forms (see e.g. Johansson, 2006 for possible skewed distributions). It is also not always possible to unambiguously distinguish between restrictive and non-restrictive relative clauses. In these instances, the ambiguous tokens should be excluded (Huber, 2017: 81).

Similarly, adverbial relatives should also be excluded. They admit a different inventory of variants, including the relative markers *where* and *when*, and are therefore not part of the same variable context as

restrictive relatives (Tottie and Rey, 1997: 229). An example of an adverbial relative from (Tagliamonte et al., 2005: 86) is given below.

- (13) And that was *the only time*  $\emptyset$  he'd even been hospital in his life.

Furthermore, as Tagliamonte (2013: 96) notes, much of the work on the relative system has either only looked at written language, or not distinguished systematically between written and spoken language (e.g. Hinrichs et al., 2015). This has serious consequences for estimates about the distribution of variants in the relative system in English. The *wh*-relatives are associated with more formal contexts, and can therefore be expected to show higher rates in the written language (Huber, 2017; Romaine, 1980: 234; Tagliamonte, 2013: 95). Not distinguishing between written and spoken language will likely lead to disproportionate estimates of the rates for *wh*-relatives (Ball, 1996: 229; Tottie & Rey, 1997: 244). For example, Johansson (2017: 270) notes an increase in *wh*-forms during the Early Modern English period, but this change only appears to take place in conventional written genres (e.g. literary texts, prose, scientific texts), and is not operative in comparable surrogates of orality (e.g. trials and drama).

One of the most robust constraints on the selection of relative markers is the grammatical role of the antecedent. Restrictive relative clauses can be divided into two types – subject relatives, and non-subject relatives. The category of non-subject relatives contains direct object relatives, genitives, and oblique relative constructions (Tagliamonte, 2013: 98).

*Whose* is restricted to genitive constructions entirely (Ball, 1996: 232). Examples of subject (14) and object (15) relatives from Tagliamonte et al. (2005: 87) are given below. In mainstream varieties of Present Day English, zero subject relatives are typically rare and are found in very specific constructions (*there*-existentials and *it*-clefts). Note the appearance of non-standard and vernacular features (non-standard *were*, regional word *wee*, meaning ‘small’ or ‘little’, and *summat* meaning ‘something’) alongside the non-standard zero subject relative in (14c), indicated by  $\emptyset$ .

- (14) a. He was asking one or two people *who* he thought might be interested.  
b. The broad words *that's* used here would be similar.

- c. And there were a wee alarm clock Ø sat on the window.
- (15)
- a. The last meeting Ø we had in that church.
  - b. And they used the old nets which we could call strabbles.
  - c. Then there were a word that I couldn't get summat to rhyme with.

The effect of the grammatical role of the antecedent can be linked to another constraint on relative selection – the animacy of the antecedent. Human antecedents are favoured in subject position. This means that relatives which are favoured with human antecedents (e.g. *who/whom*), are more likely to be selected when the antecedent is a subject (Ball, 1996: 232; Tagliamonte, 2013: 98). *Which*, on the other hand, is reported to be associated more closely with inanimate antecedents. Tagliamonte (2013: 99) reports that relative *that* is strongly favoured with inanimate antecedents, but is still more frequent with human antecedents than *who* across the contemporary varieties studied (Northern Ireland, Lowland Scotland, Northwest England, Northeast England, and Southwest England).

Tagliamonte et al. (2005) examined relativiser selection in several dialect areas: Northern Ireland (Portavogie and Cullybackey), Lowland Scotland (Cumnock), and Northwest England (Maryport). In Tagliamonte (2013), this data is analysed alongside comparable data from Northeast England (York, Wheatley Hill) and Southwestern England (Tiverton in Devon, and Wincanton in Somerset).

The type of antecedent constrains the choice of relative marker in Present Day English, albeit with conflicting claims in the literature. Huddleston (1971) associates overt relativisers with indefinite antecedents, while zero appears more with pronominal and definite antecedents. Elsewhere, zero is associated instead with indefinites in subject position (Tottie & Harvie, 2000: 209). Tagliamonte et al. (2005: 98) divided antecedents into three categories – definite NPs, indefinite NPs, and indefinite pronouns. Tagliamonte et al. (2005: 101) found that a zero relative marker was more likely to be used with an indefinite antecedent, but that definiteness was a weak effect on the selection of non-subject relatives.



The type of matrix clause has also occasionally been associated with effects on relativiser choice (Ball, 1996; Quirk et al., 1985: 416; Tagliamonte et al., 2005: 96-07). In particular, zero subject relatives are claimed to be favoured in existential constructions, cleft sentences, and possessives. Examples from Tagliamonte et al. (2005: 96) are provided below.

*Existentials:*

- (16)
- a. There's a girl fae Catrine Ø has got all medals.
  - b. There's no many folk Ø liked going to the pit to work.
  - c. Sure, there a good when Ø doesnae know our language.

*Cleft sentences:*

- (17)
- a. It's a castle Ø they live in.
  - b. It was World Cup Ø was on.
  - c. It was an earthen floor Ø was in that house.

*Possessives:*

- (18)
- a. Aye, I'd a friend Ø come down there to work.
  - b. I have a woman Ø comes in on a Thursday morning.
  - c. It's only I have these old dolls Ø aie had names.

Another constraint on the selection of relatives is the length of the relative clause (Tagliamonte et al., 2005: 97). Longer relative clauses, particularly those longer than four words, favour overt relativisers, and disfavour the zero relative. This has been connected to processing constraints: with a non-overt relative marker (e.g. the zero relative), a longer relative clause risks ambiguity and confusion. It has been claimed that more cognitively complex constructions favour explicit variants over non-explicit variants, as per Rohdenburg's (1996: 151) *Complexity Principle*. However, Tagliamonte et al. (2005: 102) found that while the zero relative was strongly favoured in shorter, more simple clauses, overt markers were not necessarily favoured in longer clauses.

A similar processing-related constraint concerns the adjacency of the relative marker and the antecedent. Again, overt markers are distinguished from the zero relative – the former are favoured in non-adjacent contexts, while the latter is primarily restricted to contexts where the relative marker is adjacent to the antecedent (Tagliamonte et al., 2005: 98). However, this effect is not selected as significant in any community examined by Tagliamonte et al. (2005: 100).

As the research reported in this dissertation is conducted in a variationist framework, it is also necessary to discuss the social conditioning of relativiser selection. Romaine (1982) claims that the *wh*-relatives are a change from above. It might therefore be reasonable to expect that the relative system would be subject to social differentiation, and may in some way reflect the complex social dynamics of the target community. This issue is taken up by Cheshire (2005: 500), who notes that the relative scarcity of syntactic variables in running discourse makes them less salient, and therefore less permeable to social evaluation. This is particularly notable in comparison to phonological variables. Similarly, Tottie and Rey (1997: 245) state that relativisation is a covert variable. This renders relativisation less susceptible to social evaluation as it operates below the level of speaker consciousness (Tottie & Rey, 1997: 245).

However, Tottie and Rey (1997: 242) did find that sex was a significant factor in relative selection in their data, with female speakers leading in the use of *that* and *what* (characteristic of the Early African American Vernacular English they were looking at), while male speakers used the zero relative more frequently. They add, however, that their small sample size, and sparse information about social roles in the community sound a note of caution in drawing definitive conclusions.

Varieties that are further from the mainstream are found to consistently have lower rates of *wh*-forms than those which are spoken in more urban centres (Tagliamonte, 2013: 99-100). Such a claim might be related to the origin of the *wh*-forms, and the idea that this was a change from above, originating in the educated upper echelons of society and subsequently diffusing down the social hierarchy (Romaine, 1982). The *wh*-forms are originally connected with more syntactically complex relativisable positions (such as oblique and genitive relative constructions; see e.g., Romaine 1982), and therefore are closely related to level of education, which may be reflective of speaker social status. Similarly, Adamson (1992: 129) found social status to significantly affect variant selection, with the lower-status speakers producing a higher proportion of zero than upper-class speakers.

Huber (2017: 84) examined the effects of both sex and social class on relativiser selection in 19<sup>th</sup> century British English in the Old Bailey Corpus. Social class was not found to be a significant factor, but sex was. Male speakers used *wh*-relatives more frequently, while female speakers predominated in the use of *that* and zero. In contrast, D’Arcy and Tagliamonte (2010: 395) found a sex effect in contemporary Toronto English, but only when speakers were divided by education. The effect was only present among speakers with post-secondary education, where female speakers used *who* more frequently than male speakers, who led in the use of *that*. Speakers without post-secondary education showed no sex effects. D’Arcy and Tagliamonte (2010: 393) also found age effects – the prestige variants (in their case, *who*) were used most frequently by speakers aged 30-59 (i.e. middle-aged), while less prestigious variants (*that* and zero) found favour with younger and older speakers.

The extent to which relativiser variation is conscripted for the purposes of marking social differentiation therefore remains an unresolved question. Social class seems to have a lesser effect on relativiser choice, and may only do so when intersecting with sex effects (Huber, 2017). Finally, D’Arcy and Tagliamonte (2010) point to the presence of age-grading effects in Toronto English as evidence that the relative system has stabilized, and that the *wh*-forms carry prestige. It will be interesting to see whether this is also the case in the English of 17<sup>th</sup> century Massachusetts, or whether contemporary patterns reported in the literature reflect more recent developments.

### 3.2.3 Relativisation in Old English

As mentioned, Present Day English distinguishes between restrictive relative clauses and non-restrictive relative clauses, with the latter disfavoured the selection of *that*. This pattern was present in Old English in a weaker form, with the indeclinable particle *þe* being favoured in restrictive relative clauses. Unless otherwise specified, all Old English examples below are from Traugott (1992).

There were three options for introducing a relative clause in Old English: a pronominal relativiser, an invariant particle, or a zero relative (Suárez, 2012: 79; Traugott, 1992). In Old English it was possible to use the demonstrative pronoun *se*, *seo*, *þæt* ‘that’ as a pronominal relativiser akin to Present Day English

*who, whom, whose*. The pronominal relativiser may also be followed by the invariant particle *þe* (Traugott, 1992: 224).

- (19) *Donne is an port on suðewearðum þæm lande (DAT),*  
 Then is one port in south-of that land,  
*þone (ACC) man hæet Sciringes heal*  
 which one calls Skiringssalr  
 ‘Then there is a port in the south of that country which is called Skiringssalr.’

- (20) ... *þæt heo ne woldon heora Gode (DAT) hyran þone (ACC) þe*  
 ... that they not wanted their God to-obey whom PT  
*heo gelyfden*  
 they believed  
 ‘... that they did not want to obey the god in whom they believed.’

The second relativisation strategy in Old English is with an invariant particle, *þe*. The invariant particle primarily appears in subject or object relatives (Traugott, 1992: 226).

- (21) *sealed þæm munucum corn genog þe wæron æt Hierusalem*  
 gave those monks corn enough who were at Jerusalem  
 ‘Gave enough corn to the monks who were in Jerusalem.’

It can also (less frequently) be used in dative (22) and genitive (23) constructions (Traugott, 1992: 226). Note that in (22), the preposition is stranded.

- (22) ... *nyhst þæm tune ðe se deada man on lið*  
 ... next that homestead PT that dead man in lies  
 ‘... next to the homestead in which the dead man lies’

- (23) ... *sio hea goodness þe he full is*  
 ... that high goodness PT he full is  
 ‘... the great goodness of which he is full.’

Invariant *þe* is typically favoured with singular antecedents that are modified by demonstratives, or with antecedents that are modified by certain quantifiers (Traugott, 1992: 226).

There are also instances of invariant *þæt*, which requires the preposition to be stranded. Cases of invariant *þæt* were rarer than invariant *þe*, but are interesting nonetheless in the context of total replacement of *þe* by *that* in Middle English (Traugott, 1992: 227). A further example of this from Molencki (1987: 95) is given in (25).

- (24) *þurh þa halgo rode (FEM ACC) þæt Crist wæs on þrowod*  
 through that holy cross that Christ was on tortured  
 ‘through the holy cross on which Christ suffered.’

- (25) Wæs he ærest ærcebiscopa, þæt him eall Ongolcyn  
 was he first archbishop to whom him the whole English race  
 hyrnesse geþafode  
 obedience yielded.  
 ‘He was the first archbishop to whom the whole English race yielded obedience.’

Zero relatives were rare in Old English (Suárez-Gómez, 2006: 36; Traugott, 1992: 228). This was particularly the case with zero object relatives, to the extent that their status as relatives has been called into question (Mustanoja, 1960: 205; Romaine, 1982: 74). In Old English the zero relative is almost entirely restricted to subject position (Levey & Hill, 2013: 34; Traugott, 1992: 228). This is in contrast to modern mainstream varieties, where, outside of *there* existentials and *it*-clefts, zero is rare in subject position, although this construction continues to be found in modern English dialects, albeit at relatively low rates (Tagliamonte et al. 2005). The zero subject relative in Old English is typically found with stative

verbs, including *wesan* ‘to be’, as in (26) below, *hatan* ‘to call, name’, *belifan* ‘to remain’, and *nyllan* ‘to not want’ (Traugott, 1992: 228).

- (26)        &    on þys ilcan gere forðferde æþered    wæs on Defenum ealdorman  
              and in this same year died        Aethered was in Devon chief  
              ‘and in this same year Æthered, [Ø] was chief of Devon, died.’

### 3.2.4 Relativisation in Middle English

Relativisation in Middle English is best characterised by two important developments. The early part of the period is exemplified by competition between two strategies of invariable relativisation – the indeclinable particle *þe*, and indeclinable *þat* (Suárez, 2012: 81). This competition is said to have its origins in northern varieties of (late 9<sup>th</sup> century) Old English (Molencki, 1987), and has been connected with the linguistic influence of the Scandinavian invasion. Areas that were less affected by these invasions, e.g. the South and Kent, retained higher rates of invariable *þe* for longer (Fischer, 2008: 296; Suárez, 2012: 101).

*Þe* is by far the dominant relative strategy in Old English, and in early Middle English (1150-1250). However, by the end of the early Middle English period, there is evidence of an increase in *þat* at the expense of *þe* (Suárez, 2012: 85). An example of relativiser *þat* in Middle English from Suárez (2012: 80) is given below in (27).

- (27)        Morgan wolde haue hade alle þe lande fram bigende Humber,  
              Morgan would have had all the land from beyond Humber  
              *þat* Conedag helde.  
              that Coneday held  
              ‘Morgan wanted to have had all the land from beyond the River which Coneday held.’

By the end of the 14<sup>th</sup> century, *that* was the sole remaining invariable relative marker, and had almost completely supplanted erstwhile pronominal demonstrative relativisers such as *se*, *seo*, and *þæt* (Fischer, 2008: 295; Fischer et al., 2000: 91; Suárez, 2012: 83). *Þe* is non-existent from the 15<sup>th</sup> century onwards

(Suárez, 2012: 84). *That* had become the relativiser of choice, regardless of the restrictiveness of the relative clause (i.e. it was dominant even in non-restrictive relative clauses, a clear difference with Present Day English), or the animacy of the antecedent (Suárez, 2012: 86).

The Middle English period also saw the rise of another major development: the use of *wh*-forms in a relativising function (Romaine 1982). The *wh*-relatives are said to have originated as interrogative pronouns, which then acquired a relative interpretation (Romaine, 1981: 80-81). Suárez (2008) connects the rise of the *wh*-relatives with the gap left by the disappearance of Old English *se*, *seo*, *þæt*, while Mustanoja (1960) claims Latin (and later French) influenced the emergence of the *wh*-relativisers. This was a gradual process – *which*, *whom*, and *whose* appear in a relativising function before *who* does (Johansson, 2017: 277).

*Which* was initially very infrequent in the 12<sup>th</sup> and 13<sup>th</sup> centuries, and only began to compete with *that* from the 14<sup>th</sup> century onwards, particularly in more formal prose, with limited success until the 15<sup>th</sup> century (Fischer et al., 2000: 92). *Whom* and *whose* were even rarer in the early Middle English period. Each of these relatives was primarily preceded by a preposition, and was found more frequently in non-restrictive relative clauses than in restrictive relative clauses (Fischer et al., 2000: 92). As the *wh*-relatives became more widely established, *that* was increasingly associated with restrictive relative clauses (Romaine, 1980: 222).

The earliest known use of *who* is in the early 15<sup>th</sup> century, with the highly specific function of referring to a deific antecedent, and in formulaic closings to letters (Fischer, 2008: 301; Romaine, 1980: 223). It retained this epistolary usage until the 16<sup>th</sup> century. The earliest known use of relative *who*, from Rydén (1983), cited in Fischer (2008: 301), is given below.

(28) I submitte me and alle þis matier to yowr good discrecion, and euere gremercy God and ye,  
**who** euere haue yow and me in his gracious gouernance.

(Paston, 1426)

The use of relative *who* completes the paradigm alongside (genitive) *whose* and (object) *whom* for personal antecedents (Fischer, 2008: 301).

Much like Old English, the zero relative in subject position is highly constrained in Middle English. It retains the association with stative verbs, such as *to be*, and often appears in existential *there*-constructions (Fischer et al., 2000: 94). This is reminiscent of zero subject relatives in vernacular Present Day English. Examples of Middle English zero subject relatives are given below (Fischer et al., 2000: 94).

- (29) I know no knight in this contrey [ø] is able to macche hym.  
'I know no knight in this country [who] is able to match him.'  
(Malory *Works* 377.36-6)

- (30) Adam ben king and eue quuen/ Of alle ðe ðinge [ø] in werlde ben.  
Adam are king and Eve queen of all the things in world are  
'Adam and Eve are king and queen of all the things that are in the world.'  
(*Gen. & Ex.* 297-297)

A major difference between Old English and Middle English is the gradual appearance of zero object relatives. These were limited to very specific contexts, ultimately appearing with one of two verbs – *clepen* 'call', or possessive *have*, reminiscent of the stative verb constraint on zero subject relatives. Examples of zero object relatives in Middle English from Fischer et al. (2000: 94) are given below.

- (31) Of Northfolk was this Reve of which I tell,/ Beside a toun [Ø] men clepen Baldeswelle  
'From Norfolk was this Reeve I am telling you about, close to a town people call Baldeswelle.'
- (32) Sir, be þe faith [Ø] i haue to yow...  
'Sir, by the faith I have to you...'



The Middle English period also witnesses the appearance of a preposition stranding option for *wh*-relatives. While preposition stranding is (somewhat controversially) available with Old English relativisers, in the early Middle English period *wh*-relatives were obligatorily pied-piped (Bergh & Seppänen, 2000: 302). The first recorded instance of stranding with a *wh*-relative is in the 13<sup>th</sup> century. Contemporaneous data finds very low rates of stranded prepositions (Bergh & Seppänen, 2000: 302) – though as the data is drawn from theological and literary sources, this is not surprising and may not accurately reflect the vernacular situation of the time. In the Early Modern English period, these rates increase to around 11% stranding with *wh*-relatives (Bergh & Seppänen, 2000: 302). This strongly suggests an expansion of stranding with *wh*-relativisation.

The appearance of *wh*-relatives was a gradual process. According to Romaine (1980: 225), the *wh*-relatives appeared first in the most complex written styles. In the literature, this effect is attributed to a hierarchy of complexity such as that proposed by Keenan and Comrie (1977). This is a cross-linguistic approach to establishing the accessibility of specific syntactic positions to relativisation (Keenan & Comrie, 1977: 66). Essentially, certain positions are more accessible to relativisation, and this hierarchy is claimed to be consistent across languages. Less accessible positions (i.e. those lower down on the hierarchy) are associated with increased complexity (Keenan & Comrie, 1977: 88). Subjects are the most accessible, followed by direct objects, indirect objects, obliques, genitives, and finally objects of comparison. If a language is able to relativise an indirect object, then it is presumed to also be able to relativise direct objects and subjects (Keenan & Comrie, 1977: 66).

The logic here is that *wh*-relatives originated in the least accessible, most complex positions (genitives and obliques), and in the most complex styles (e.g. formal written language). It has already been mentioned that these are the contexts where *wh*-pronouns first arise – *whose* in the genitive, and *whom* and *which* in prepositional contexts (Fischer et al., 2000: 92). From here, they diffused to other syntactic positions, later becoming available in the more accessible positions and in the spoken language, although the extent to which this occurred, following Romaine (1982: 212), remains in dispute. In the history of the English relative system, then, *wh*-forms represent a new relativisation strategy that is initially (and continues to be) strongly indicative of more formal styles (D’Arcy & Tagliamonte, 2010: 387; Huber, 2017: 75).



The relativiser system in Early Modern English has received explicit treatment in several studies (e.g. Ball, 1996; Dekeyser, 1984; Dekeyser & Ingels, 1988; Johansson, 2012, 2017; Tottie & Johansson, 2015). The results from two studies, Ball (1996) and Johansson (2017) will be discussed here to highlight the key characteristics of the relativiser system in this time period.

The speech surrogate data in Ball (1996) is drawn from two sources, representing British and American English respectively: the *British State Trials*, and the *Salem Witchcraft Papers*. The *Salem Witchcraft Papers* data is restricted to personal subject restrictive relatives, while the State Trials data contain both personal and non-personal subject restrictive relative clauses. Non-subject relatives are not analysed. Consequently, any discussion of Ball (1996) below refers explicitly to subject relatives only.

Ball (1996: 246) finds *which* to be well established with non-human antecedents in 16<sup>th</sup> century written British English data, drawn from a corpus of fiction, treatises, and essays (Ball, 1994). *Which* is also present with human antecedents, but the data shows a slight preference for *who* in these contexts (11% compared to 9% *which*). Even in the written language, *that* predominates in both personal and nonpersonal subject restrictive relative clauses with rates of 80% and 65% respectively. The dominance of *that* is slightly less in non-personal contexts, while *which* is stronger (33%). Zero is absent from personal subject restrictive relative clauses in this written data, and only marginally present in non-personal contexts. As this data is based on written English, it has limited comparability with other speech-based sources. Nonetheless, it is still interesting to note the high rates of *that*, even in writing.

Turning now to Ball's (1996: 247) 17<sup>th</sup> century data, there are some stark differences. Beginning with British English prose data (treatises, essays, and fiction), it seems clear that the *wh*-relatives have made inroads. With personal subject restrictive relative clauses, there are rates of 57% for *that*, compared to 42% for *who*. *Which* is negligible at 1%, and zero is entirely absent. For non-personal subject restrictive relative clauses, the situation is somewhat reversed. *Who* is absent, and zero is negligible at 1%. *That* is the preferred form, accounting for 74% of written tokens, while *which* is responsible for 26%. Ball (1996: 246-247) argues that these results reflect several changes in the language, including the reinforcement of an animacy constraint between *who* and *which* (restricting them to human and non-human antecedents

respectively), and the increasing relegation of *that* to nonpersonal subjects. These changes, Ball (1996: 247) claims, are the basis for the Present Day English dominance of *who* with personal subject restrictive relative clauses.

In 17<sup>th</sup> century trial-based data, *that* is the predominant relative with both human and non-human antecedents by a considerable margin. In the British data, *which* is effectively absent from personal subject restrictive relative clauses. *Who* is better represented in State Trials 1 (ST1) (13%) than in State Trials 2 (ST2) (1%).<sup>6</sup> A similar effect is found in non-personal subject restrictive relative clauses with *which*. This represents 10% of tokens in ST1 compared to 4% in ST2. These results may be indicative of the class and education make-up of the speakers in each set, but a more detailed breakdown of each speaker group would be needed to make any concrete claims to this effect. Interestingly, the zero relative is more frequent among the speakers in ST2, perhaps indicating its association with less prestigious speakers and contexts. Coupled with the higher rates of *wh*-forms with ST1 speakers, this could lend support to Romaine's (1982) claim that *wh*-relatives are a change from above.

Ball (1996) also analysed a subset of the *Salem Witchcraft Papers*, containing only personal subject restrictive relative clauses, as an approximation of 17<sup>th</sup> century spoken American English. In this dataset, *that* remains the relativiser of choice (77%). *Who* is rare (4%), while personal *which* is more frequent at 11%. The combined *wh*-rates of 15% are in line with those of ST1. However, as Ball (1996: 248) notes, these higher rates of *wh*-relatives do not reflect a cross-Atlantic difference, but instead reflect high rates of *which* in the formulaic context *afflicted by a woman which told me her name was X*. As such, the underlying rates are likely much closer to those for ST2.

Ball (1996: 252) claims that the development of the person constraint, whereby *which* is restricted to non-personal antecedents and *who* comes to dominate personal subject contexts, takes place during the 17<sup>th</sup> century. This change has its origin in the written language (Ball, 1996: 247-248), and does not yet appear in Ball's (1996) speech-like data. If this is the case, then it might be expected that data from the *Salem*

---

<sup>6</sup> Ball (1996: 246) divides the State Trials data into two separate groups based on a loose approximation of social class – State Trials 1 contains data from lawyers and those with titles, while State Trials 2 holds data from all other speakers.

*Witchcraft Papers* in 1692 will not yet demonstrate this pattern, particularly in the more speech-like documents (e.g. examinations).

Johansson (2017) examines the use of relativisers in Early Modern English in the Corpus of English Dialogues 1560-1760 (CED). Data were drawn from trials and dramas spanning two time periods: 1560-1599, and 1680-1719. In terms of overall distribution, *that* predominates in both time periods (around 50%), showing greater frequencies than the *wh*-forms combined (around 35%). Zero is more frequent than reported in Ball (1996) with rates of 11% in the earlier period and an increase to 21% in the later period (Johansson, 2017: 269). This near-doubling in the rates for zero appears to represent a particularly striking finding. However, these results are from a dataset combining restrictive and non-restrictive relative clauses. When Johansson (2017: 271-273) separates the two disparate types, the rates for *wh*-forms collectively drops to 21%, in line with the 21% of zero. The proportion of *that* is higher when non-restrictive relative clauses are removed, rising to 58% of all tokens. *Which* is the most frequent out of all *wh*-relatives (11%), while *who*, *whom* and *whose* are all low (5%, 2%, and 3% respectively).

Johansson (2017: 274) also considers the impact of two major constraints – the grammatical role of the relativiser, and the animacy of the antecedent. Unfortunately, it is not possible to track changes in the relative system with this data, as it is not separated by time period. In subject relatives, *that* is the relativiser of choice, regardless of animacy (69% with personal antecedents compared to 62% with non-personal antecedents). The zero relativiser is predictably rare in subject position, and does not show animacy effects (3% for both human and non-human antecedents). The primary distinction in terms of animacy is among the *wh*-relatives. With personal antecedents, *who* accounts for 24% of tokens, while *which* accounts for 2%. *Whom* is unsurprisingly low at 1%. With non-personal antecedents, the ranking is reversed. *Which* is more frequent than *who* (34% compared to 1% respectively).

In non-subject contexts, the picture is strikingly different. With personal antecedents, *whom* is the relativiser of choice at 41%, followed by zero with 31%. *That* is responsible for 25% of personal object contexts. *Which* is clearly disfavoured here with rates of only 3%. In object relative constructions with non-personal antecedents, the relative system is split almost equally among three variants – *that*, *which*, and zero (31%, 33%, and 36% respectively).

It seems fairly clear based on this data that Ball's (1996: 246-247) claim that *that* was confined to non-personal antecedents was not operative in written 17<sup>th</sup> century English approximations of speech. The rates for *that* with personal and non-personal antecedents in the data are almost equivalent, both in subject and object constructions (Johansson, 2017: 274).

Based on the discussion above of the data provided in Ball (1996) and Johansson (2017), several key generalisations can be made. In both studies, *that* is the predominant relative marker in almost all contexts. The sole exception is with non-personal object constructions (Johansson, 2017: 274). The initial stages of *which* being increasingly confined to non-personal antecedents can be observed in Johansson's (2017) 17<sup>th</sup> century speech-related data. The animacy constraint proposed by Ball (1996) appears to already be present, with *who/whom* predominating with human antecedents, and *which* being the *wh*-relative of choice with non-human antecedents.

There also appears to be a quite drastic decrease in rates for subject zero in this period. Tottie and Johansson (2015) examine the distribution of zero subject relatives in Early Modern English, again using the *Corpus of English Dialogues*. The data span five 40-year periods, from 1560-1760. The results demonstrated a gradual decline in rates of zero subject relatives, particularly in trial-based data. The zero relative was found to be favoured when the relative marker and the antecedent are adjacent. Zero subject relatives in Early Modern English are primarily restricted to specific constructions – specifically *there*-existentials. An example from Tottie and Johansson (2015: 142) has been reproduced below.

(34)        There was a Gentleman Ø fetch'd Me and the Clerk from the Fleet.

The rates over time reported in Tottie and Johansson (2015) for zero subject relatives, when compared with those for zero subject relatives in Johansson (2017), suggest that zero was becoming more closely associated with non-subject relativisation. This is a reversal of the situation in Middle English – recall that in Middle English, the zero relative was favoured in subject position more than object position.

Summarizing, the relative system in English has undergone substantial changes in the past 1500 years – the disappearance of Old English *þe*; the replacement of *se*, *seo* and *þæt* by *that* in Middle English; the rise of the *wh*-relatives throughout the Middle English period; and in Early Modern English, the restriction of relative *which* to non-personal contexts and the increase in zero relatives in non-subject position. The inventory of relatives available in Present Day English have their origin in Early Modern English, but continue to be subject to change at the level of constraints governing their selection. While relativisation in English is a well-studied topic, it is by no means exhausted, and methodological and data-based issues ensure that there is scope for further research.

The current chapter will seek to address the key determinants of relative marker choice in 17<sup>th</sup> century English. In so doing, there are several key methodological issues that need to be addressed at the outset. A first important issue includes the separation of restrictive and non-restrictive relative clauses, and subject from non-subject relatives. A further requirement involves distinguishing data from different genres that represent the written register, and those that more closely approximate spoken language. Also needed is access to a substantial amount of data to enable reliable quantitative analysis as well as, where possible, the availability of detailed sociobiographical information about the speakers who produced the data as well as information concerning the community that they lived in.

A study of relativisation in the *Salem Witchcraft Papers* is able to meet these requirements. The *Salem Witchcraft Papers* provide a substantial amount of data that can be fairly neatly categorised in terms of speech-likeness: warrants and indictments are more formulaic and routinised, and are less likely to accurately represent speech of the time; examinations may be reasonably assumed to approximate 17th century spoken American English; and witness depositions and testimonies occupy a middle ground.

The inventory of available relative markers had become established by the early 18th century, but the distributional rates for relative markers would be susceptible to change (Dekeyser, 1984: 61; Huber, 2017: 75-76). It will therefore be valuable to investigate whether 17th century American English is organized in terms of a tripartite system of relativisation like Present Day English. Building on Romaine's (1982) claims, it will be interesting to consider how far the *wh*-relatives have infiltrated this historical variety of spoken English, and how well this system resembles the distribution of variants in the relative system as

reported in the literature (Ball, 1996; Dekeyser, 1984; Dekeyser & Ingels, 1988; Johansson, 2012, 2017; Romaine, 1982; Tottie & Johansson, 2015).

The analysis will contribute to the body of empirically accountable, quantitative studies of the relative system in earlier varieties of English, offering a comparable and replicable approach, and specifically providing insight into the language used in 17<sup>th</sup> century Massachusetts.

### 3.3 Methodology

#### 3.3.1 Variable Context

Adhering to the Principle of Accountability (Labov, 1972: 72), I systematically identified and extracted every available instance of a restrictive relative clause, along with its surrounding context (roughly 5-10 words either side of the relative clause). The presence of a null variant necessitated manual extraction of tokens through close reading, as an automated extraction process would not allow identification of zero relatives. Including the zero relative in the data is essential to satisfy the Principle of Accountability (Labov, 1972: 72). In order to provide a dataset that is empirically aligned with previous research on relativisation, non-restrictive relative clauses have been excluded from this study. The strong association of non-restrictive relative clauses with *wh*-relatives motivates their separation from restrictive relative clauses. I define a restrictive relative clause as one which restricts the referent to a specific subset, while a non-restrictive relative clause provides additional information (Tagliamonte et al., 2005: 85). Non-restrictive relatives, which show very little variability and primarily appear with *wh*-relatives, were excluded from the dataset. Relevant tokens were identified through close reading of each text.

*Restrictive relative clauses:*

(35) is y<sup><D></sup> y<sup>e</sup> Same man y<sup>t</sup> appeared before to you?

(SWP125.4 – John Hawthorne,  
Examination: 1<sup>st</sup> March 1692)

(36) he brought to me a new fashion book ~~to me~~ **which** he did not vse to bring

(SWP022.21 – Mercy Lewis,  
Deposition: 3<sup>rd</sup> August 1692)



*Non-restrictive relative clauses:*

- (37) This Deponent hath mett with severall other of her Pranks. at severall times: **which** would take up a great time to tell of

(SWP013.8 – William Stacy,  
Deposition: 30<sup>th</sup> May 1692)

- (38) Immediatly there appeared to me mr. Burroughs two first wiues ~~w<?>~~ in their winding sheets **who<m>** I formerly well knew

(SWP022.27 – Mary Walcott,  
Deposition: 3<sup>rd</sup> August 1692)

Several constructions which do not belong to the target variable system were excluded from the data. For example, adverbial relatives are excluded from this study on the basis that they fall outside of the envelope of variation; they pattern differently to other types of relative clause, and allow a different set of variants (including e.g. *when*, *where*, *why*). This includes temporal relatives, as below.

- (39) exept about that time nor since that time exept by ackcident **that** it was hurt

(SWP049.5 – William Beale,  
Testimony: 8<sup>th</sup> August 1692)

- (40) her Mistris: told her she had set her hand to y<sup>e</sup> devils book: that same night: **that**: I was thrown out of bed

(SWP135.3 – Mary Warren,  
Examination: 21<sup>st</sup> April 1692)

Tokens that were ambiguous (for several reasons) were excluded from the data. (41) was excluded as it may be read as either a relative clause or a complementizer clause.

(41) Marey Lacey what tricks doe You know Ø hee has Done

(SWP087.2 – unknown speaker,  
Examination: 21<sup>st</sup> July 1692)

### 3.3.2 Internal Factor Groups

Tokens were coded for a number of factor groups (both internal and external) abstracted from previous investigations of relativisation.

#### 3.3.2.1 Syntactic Function of the Relative Marker

Each relative clause was coded for the syntactic function of the relative marker, a key determinant of variant selection (Ball, 1996: 230). Subject relatives, non-subject relatives (object and oblique relatives) and genitives were coded separately. Zero object relatives have been claimed (e.g. Tottie & Johansson, 2015) to increase in frequency in the Early Modern Period. Zero subject relatives, on the other hand, were more frequent in Middle English, but declined sharply in Early Modern English (Fischer, 2008; Fischer et al., 2000).

#### *Subject relative*

(42) I <sup>^</sup>{we} oftē spake to hur of some things y<sup>t</sup> wer reported of hur

(SWP072.9 – Daniel Warner Sr.,  
Statement: 25<sup>th</sup> June 1692)

#### *Object relative*

(43) it should Cost his father Zachariah Herick one; or two of y<sup>e</sup> best Cowes **which** he had

(SWP063.23 – Sarah Good,  
Deposition: 29<sup>th</sup> June 1692)

#### *Oblique relative*

(44) in all this time nothing has been made Appear for w<sup>ch</sup> shee deserv<sup>d</sup> Imprisonment or death

(SWP115.2 – Nicholas Rice,

*Genitive relative*

- (45) there was a minister a litle man **whose** name is Burroughs

(SWP128.2 – Mary Toothaker,

Examination: 30<sup>th</sup> July 1692)

*3.3.2.2 Animacy*

As mentioned earlier, one of the defining features of the relative system in Early Modern English, as claimed by Ball (1996), is the emergence of an animacy parameter. In this period, it is proposed that *which* becomes confined to non-human antecedents. Similarly, *that* has been proposed to also become increasingly associated with non-human antecedents, though Ball (1996) points out that by the end of the 17<sup>th</sup> century, this shift is primarily literary and is absent from speech-like data. In order to test whether this animacy constraint is operative in my speech-like data, tokens were coded based on a binary distinction between human and non-human antecedents.

*Human antecedent*

- (46) they saw the Apperishtion of the other woman **that** appered with gooddy estick the othr day

(SWP045.17 – Samuel Abby,

Deposition: 23<sup>rd</sup> May 1692)

*Non-human antecedent*

- (47) To help take downe y<sup>e</sup> Celler wall of The owld house Ø she formerly Liued in

(SWP013.12 – John Bly Sr.,

Testimony: 2<sup>nd</sup> June 1692)

*3.3.2.3 Matrix Clause Type*

Type of matrix clause is noted in the literature as an important factor in the selection of the zero relative (e.g. Levey & Hill, 2013; Tottie and Johansson, 2015). Specifically, the appearance of zero subject

relatives, though rare, has been connected with *it*-clefts and *there*-existentials. Clefts and existentials receive variable treatment in the literature, with controversy over their status as relative clauses. Some studies on relativisation include both (e.g. Levey & Hill, 2013; Tagliamonte et al., 2005), some exclude both (e.g. Huber, 2017), and others include one or the other (e.g. Ball, 1996). Due to this factor group's claimed importance in the selection of zero subject relatives, I have retained main clauses comprising clefts and existential-*there* constructions in the present study, with the prediction that these constructions will favour (subject) zero relatives. Stative possessives were also coded as a separate factor, as this matrix clause type has been associated with subject zero relatives (Tagliamonte et al., 2005). All other matrix clauses were coded as 'other'.

#### *Existential*

- (48) She said she would confess <if> she could But that there was something at her breast **that** hindered her.

(SWP128.2 – Mary Toothaker,  
Examination: 30<sup>th</sup> July 1692)

#### *Cleft*

- (49) she Answered it was the Enemy Ø hurt her.

(SWP122.2 – Abigail Somes,  
Examination: 13<sup>th</sup> May 1692)

#### *Stative possessive*

- (50) to y<sup>e</sup> best of my remembrance I had a Cow w<sup>ch</sup> vsed to giue a good Mess of ^{milke} twice a day

(SWP024.9 – John Rogers,  
Deposition: 2<sup>nd</sup> August 1692)

#### *3.3.2.4 Processing Constraints*

Tokens were coded for two factor groups related to processing constraints: the length of the relative clause, and adjacency of the relative marker and the antecedent. More complex constructions (i.e. longer

relative clauses or tokens where the relative marker and the antecedent are non-adjacent) are claimed to favour overt relative markers, while less complex constructions favour the zero relative (Rohdenburg, 1996). Subject relative clauses were classified as either short (1-5 words), medium (6-9 words) or long (10+). The scale for non-subject relative clauses is necessarily different, as they contain a minimum of 2 words (i.e. a subject and a verb). Short non-subject relative clauses are therefore 2-5 words in length. Tagliamonte, Smith, and Lawrence (2005: 97) note that clause length does not necessarily directly correlate with increased processing constraints, and that relative clause complexity (as measured by the presence of embedded clauses) may play more of a role in the use of relative markers. The effect of clausal complexity was not examined here, but would be an interesting angle for future research.

### *Short*

- (51) he heard a trumpet sounded at the meeting and thinks it was Burse **that** did it  
 (SWP009.2 – William Barker Sr.,  
 Examination: 29<sup>th</sup> August 1692)

### *Medium*

- (52) his house hath been troubled w<sup>th</sup> Cats & dogs w<sup>ch</sup> he saw often running up & down  
 (SWP034.9 – John Cole,  
 Deposition: 3<sup>rd</sup> October 1692)

### *Long*

- (53) o<sup>r</sup> Eldest Child **who** promised as much health & vnderstand{ing} by Countenance and actions as any other Children of his years  
 (SWP013.10 – Samuel Shattuck,  
 Testimony: 2<sup>nd</sup> June 1692)

In terms of adjacency, tokens were divided between immediate adjacency (no intervening words between the antecedent NP and the relative marker) and non-adjacency (at least one intervening word).

### *Adjacent*

- (54) 5 afflicted persons doe charge this woman to be the very ~~w~~ <wo>man **that** hurts them  
(SWP013.2 – Ezekiel Cheever,  
Examination: 19<sup>th</sup> April 1692)

*Non-adjacent*

- (55) he did syne the devils book with blood brought to him in a thing lyke an Inkhorn **that** he  
dipt his finger y<sup>r</sup>in  
(SWP009.2 – William Barker Sr.,  
Examination: 29<sup>th</sup> August 1692)

I hypothesise that processing constraints will favour the selection of overt and explicit relative markers in more complex constructions, specifically in longer clauses and when the relative and the antecedent are non-adjacent. Zero will contrastively be favoured in less complex constructions.

*3.3.2.5 Antecedent Type and Definiteness*

Literature on the effects of antecedent type and definiteness of the antecedent is inconsistent in the conclusions that are reached. In particular, the zero relative has variously been associated with pronominal antecedents (Huddleston, 1971; Piercy et al., 2011; Quirk, 1957; Tottie, 1995; Tottie & Rey, 1997), with indefinite noun phrase antecedents (Tagliamonte et al., 2005; Tottie & Harvie, 2000), or with definite antecedents (Wasow et al., 2011). Harvie (1998) finds no difference between pronouns and indefinite NPs in the selection of the zero relative. Jaeger and Wasow (2005), following Ariel (1990), propose an accessibility scale as follows: Pronoun > Demonstrative > First Name > Definite NP > Indefinite NP. Antecedents that are higher on the scale (i.e. pronouns) are considered more accessible, and therefore may be more likely to select the zero relative marker. Given the clear lack of consensus about the effect of antecedent type and definiteness on the relative system, tokens were coded for each separately. For antecedent type I distinguish between pronominal and NP antecedents.

*Pronoun*

- (56) it was she **that** made my ffathers sheep to run a[Lost]ll they ware all lost  
(SWP016.17 – Ann Putnam Jr.,

Deposition: 9<sup>th</sup> September 1692)

*Noun phrase*

(57) he saw the shape of a black dog **which** looked very fiercely upon him

(SWP010.1 – William Barker Jr.,

Examination: 1<sup>st</sup> September 1692)

For definiteness, the division is made between definite and indefinite antecedents.

*Definite*

(58) and y<sup>t</sup> arme with **which** I strook was presently disenabled

(SWP013.11 – John Louder,

Testimony: 2<sup>nd</sup> June 1692)

*Indefinite*

(59) abegetal Wiluams sd that there was a lettell black menester **that** Liued at Casko bay

(SWP022.24 – Abigail Williams,

Testimony: 22<sup>nd</sup> April 1692)

### 3.3.3 External Factor Groups

Tokens were additionally coded for several social factor groups, including sex, birthplace, age, document type, and an individual's position in the social hierarchy, in order to investigate the possible influence of social characteristics on the selection of variants.

#### 3.3.3.1 *Speaker Sex*

It has previously been claimed by Huber (2017) that female speakers (in 18<sup>th</sup> century British trial data) favour the use of the zero relative, while male speakers lead in the use of the *wh*-relatives. This finding is at odds with the notion that female speakers typically spearhead language change (Labov, 1990), and so is worth investigating.

#### 3.3.3.2 *Speaker Birthplace*

One of the main claims in Ball (1996) is that the relative system in English lacks a vernacular norm. In order to test this, speakers are grouped by birthplace, where this information has survived. Of the speakers in the data who used relativisers (and whose birthplace has survived in the historical record), 78% (N=98/126) were born in America, compared to 21% (N=21/126) born in Britain. Also present in the data are the slaves Tituba and Candy, who are thought to have been born in South America and Barbados respectively (Smith, 2018; Rissanen, 1997: 188).

#### 3.3.3.3 *Speaker Age*

The distribution of relativisers across age groups was used as an apparent-time measure of language change in progress, under the assumption that younger speakers will use the *wh*-relatives more frequently than older speakers, who will be the most frequent users of the zero relative (Huber, 2017; Tottie and Johansson, 2015: 140). To test this, speakers were divided into three age groups – younger (8-29), middle-aged (30-49), and older (50+).

#### 3.3.3.4 *Speaker Social Status*

In an attempt to avoid imposing modern perspectives of social class onto the past, the available information was used to group speakers into a loose social hierarchy, based on several key indices (e.g., wealth, education, and political power) that have been connected with social capital in 17<sup>th</sup> century Massachusetts (Zemsky, 1969). This composite metric is firmly cemented in the cultural context – the time-period, geographical area, racial and religious make-up of the population (Kerswill, 2009: 358; Mallinson, 2007: 151). There is much work that has been done on the social structure of the speech community in 17<sup>th</sup> century Salem (see e.g. Grund, 2007; Kahlas-Tarkka, 2012; Laskaris, 2019).

Any created social class hierarchy is by its very nature somewhat subjective. This means that it cannot be taken to represent an objective measure (like age), but it is nonetheless possible to use several indices to determine a speaker's position in the social hierarchy (Mallinson, 2007: 150). The rigid social classes of the feudal period are no longer as relevant – those bonds began to disappear in English society after the Black Death. Nor are the modern concepts of upper, middle, and working class entirely appropriate – these terms designate a post-industrial society (Nevalainen, 1999: 507). Instead, speakers are assigned a position



in the social hierarchy on the basis of several factors. Factors such as education, wealth, land and property ownership, and occupation were used (where available) to place speakers into a hierarchy.

Using a composite index circumvents certain issues, particularly regarding sex disparities. In the Early Modern period, the ability to read and write was not equally distributed – it was typically (at least in Western contexts) limited to the upper classes and professional men (Auer et al., 2015: 6; Nevalainen, 1999: 507). Classifying speakers solely by education level in this period would therefore unfairly favour male speakers over female (Nevalainen, 1999: 514). Issues of ownership, wealth, and occupation are also relevant – female speakers in this time period typically engaged in unpaid domestic labour, and were restricted access to wealth and resources. Interestingly (and perhaps relatedly), many of the accused were women who acted against these restrictions.

Detailed information about many of the speakers in the *Salem Witchcraft Papers* is very limited. Based on an original set of 8 divisions, speakers have been grouped into one of 4 categories: Group A, consisting of members of the gentry, clergy, and educated officials; Group B, which includes merchants and those with a specific trade; Group C, which represents those who were unspecified labourers, farmhands, and servants; and finally Group D, encompassing those speakers who were slaves. This system is intended to divide speakers into categories that broadly represent their position in the social hierarchy in 17<sup>th</sup> century colonial Massachusetts. Thus, those who owned land (typically members of Groups A and B) rank higher than those who work the land (Group C). As persons lacking their own freedom and rights, the slaves of Group D occupy the least prestigious rank in society.

Zemsky (1969: 519) identifies three indicators of social power and prestige in 17<sup>th</sup> century Massachusetts – wealth and land ownership, the inherited social standing of a famous name, and education (access to which was more accessible to those from wealthy backgrounds and the gentry). According to Zemsky (1969), these indicators played a notable role in an individual's ascension in the political sphere, particularly in the case of leaders in the Massachusetts assembly. The grouping of my speakers into a social hierarchy is based on these claims – speakers in Group A have access to all three indicators, Group B to one or two indicators (minimally wealth, but also in some cases education), while those in Group C typically did not have access to any of the above indicators of social standing. It should be noted here that

this is of course a somewhat arbitrary division. It would perhaps best be described as an approximation of the way that speakers might have been viewed by others in their society.

### 3.3.3.5 Document Type

Texts were coded as one of four document types, representing a spectrum of formality and proximity to speech. At the least formal/most speech-like end are the examinations, followed by testimonies and depositions, personal letters (including pleas and petitions) (Kytö & Walker, 2003). Official documents (such as indictments, recognizances, mittimuses, warrants) are the most formal text type in the data (Rissanen, 2003: 88). However, after the removal of set phrases (3.4.1.1) from the data, only 9 tokens from official documents remained. These tokens were retained for distributional analysis, but excluded from statistical analysis.

## 3.4 Results

### 3.4.1 Overall Distribution

Table 1 shows the overall distribution of variants in the *Salem Witchcraft Papers* data. Formulaic phrases have been excluded from this dataset (see 3.4.1.1 below). As will be shown, many of them pattern in predictable and consistent ways, and do not demonstrate any variability. Two of the set phrases (*a person REL tould me* and *the oath REL they had taken*) did in fact exhibit variability, and so have been included in the data.

| Variant                   | %  | N   |
|---------------------------|----|-----|
| <i>that</i>               | 55 | 410 |
| zero                      | 20 | 147 |
| <i>which</i>              | 19 | 138 |
| <i>who</i>                | 4  | 31  |
| <i>whom</i>               | 1  | 4   |
| <i>whose</i> <sup>7</sup> | 1  | 5   |

---

<sup>7</sup> Due to low rates and lack of variability, the genitive marker *whose* will not be discussed further. In later tables, *whose* will be subsumed under the ‘other’ category.

|                    |                         |            |
|--------------------|-------------------------|------------|
| Other <sup>8</sup> | 1                       | 8          |
| <b>Total</b>       | <b>*101<sup>9</sup></b> | <b>743</b> |

Table 1: Overall distribution of variants

*That* is the predominant relativiser in this time period by a notable margin, accounting for 55% of all restrictive relatives in the data. *Which* and the *zero* relative are almost equally frequent, with rates of 19% and 20% respectively. *Which* is the most frequently used *wh*-relative. *Who*, *whom*, and *whose* are rare, with *who* being more frequent (4%) than *whom* and *whose* (1% each). These findings mirror rather neatly Johansson's (2017: 274) British English results from the Corpus of English Dialogues (1560-1760), where *that* has rates of around 50%, *which* and *zero* are equal with 21%, and *who*, *whom*, and *whose* each represent less than 5% of the data.

In isolation, while interesting, these results do not tell us much about the selection of variants in different contexts. In order to elucidate this information, I will turn to the contribution of various factor groups to the selection of relativisers. First, however, attention will be given to the excluded formulaic phrases.

#### 3.4.1.1 Set Phrases

Eight categories of set phrase were identified in the data.

| <b>Variant</b> | <b>%</b> | <b>N</b> |
|----------------|----------|----------|
| <i>which</i>   | 41       | 31       |
| <i>that</i>    | 36       | 27       |
| <i>whose</i>   | 13       | 10       |
| <i>zero</i>    | 7        | 5        |
| <i>whom</i>    | 3        | 2        |
| <i>who</i>     | 0        | 0        |
| <i>what</i>    | 0        | 0        |

<sup>8</sup> The 'other' category is a conflation of three extremely infrequent variants – non-standard *as* and *what*, and compound relatives like *whereby*. Due to their scarcity in the data, these variants were combined into a single category.

<sup>9</sup> Percentages do not add up to 100% due to rounding.

|                               |            |           |
|-------------------------------|------------|-----------|
| <i>as</i>                     | 0          | 0         |
| <i>Compound</i> <sup>10</sup> | 0          | 0         |
| <b>Total</b>                  | <b>100</b> | <b>75</b> |

Table 2: Overall distribution of variants in set phrases

As can be seen, the overall distribution of variants in set phrases shown in Table 2 above does not match that of vernacular speech-like data. In formulaic language, *which* is the preferred form (41%), closely followed by *that* (36%). The non-standard variants *what* and *as* are absent from the set phrase data, as is any use of compound relatives and relative *who*. *Zero* and *whom* are only rarely used in set phrases (7% and 3% respectively).

Closer examination of the data reveals eight broad categories of set phrases, including legal formulae (a-e), epistolary set phrases (f-g), and direct quotes from scripture (h). These are as follows:

- (60)
- a) all that is above written
  - b) the oath which they had taken
  - c) that which by the sd Court shall be then & there Injoynd them
  - d) a person that tould me/ the same person that tould me
  - e) the felony for w'ch she stood indicted in
  - f) we whose names are underwritten
  - g) to all whome it may concern
  - h) to forgive us our trespasses as we forgive them that trespass against us

| Set Phrases                          | <i>whom</i> |   | <i>whose</i> |   | <i>which</i> |   | <i>that</i> |   | zero |   | <b>Total</b> |           |
|--------------------------------------|-------------|---|--------------|---|--------------|---|-------------|---|------|---|--------------|-----------|
|                                      | %           | N | %            | N | %            | N | %           | N | %    | N | %            | N         |
| <i>all that is above written</i>     | 0           | 0 | 0            | 0 | 0            | 0 | 100         | 6 | 0    | 0 | <b>8</b>     | <b>6</b>  |
| <i>the oath which they had taken</i> | 0           | 0 | 0            | 0 | 54           | 7 | 8           | 1 | 38   | 5 | <b>17</b>    | <b>13</b> |

<sup>10</sup> Examples of compound relatives include *whereby* and *whereof*.

|                                            |          |          |           |           |           |           |           |           |          |          |            |           |
|--------------------------------------------|----------|----------|-----------|-----------|-----------|-----------|-----------|-----------|----------|----------|------------|-----------|
| <i>that which by the sd Court</i>          | 0        | 0        | 0         | 0         | 100       | 9         | 0         | 0         | 0        | 0        | 12         | 9         |
| <i>a person that tould me</i>              | 0        | 0        | 0         | 0         | 39        | 12        | 61        | 19        | 0        | 0        | 41         | 31        |
| <i>the felony for w'ch</i>                 | 0        | 0        | 0         | 0         | 100       | 2         | 0         | 0         | 0        | 0        | 3          | 2         |
| <i>we whose names are<br/>underwritten</i> | 0        | 0        | 100       | 10        | 0         | 0         | 0         | 0         | 0        | 0        | 13         | 10        |
| <i>to all whom it may concerne</i>         | 100      | 2        | 0         | 0         | 0         | 0         | 0         | 0         | 0        | 0        | 3          | 2         |
| Scripture                                  | 0        | 0        | 0         | 0         | 50        | 1         | 50        | 1         | 0        | 0        | 3          | 2         |
| <b>Total</b>                               | <b>3</b> | <b>2</b> | <b>13</b> | <b>10</b> | <b>41</b> | <b>31</b> | <b>36</b> | <b>27</b> | <b>7</b> | <b>5</b> | <b>100</b> | <b>75</b> |

Table 3: Distribution of set phrases

While there is a limited amount of data for set phrases, the breakdown of variants by type of set phrase demonstrates variation in certain types of set phrase.

The set phrase *all that is above written*, while only yielding six tokens, categorically selects *that*. Furthermore, the allusion to written language firmly establishes these phrases as written addenda to the speech act being recorded. The phrase *that which by the sd Court shall be then & there Injoynd them* is a formulaic phrase enshrined as part of the legal process, consistently appearing in recognizances (formalised legal documents that commit the accused to appear at trial or forfeit a large sum of money), and categorically selecting *which*. The phrase *we whose names are underwritten* is formulaic and common particularly in petitions. The genitive relative *whose* was categorically selected in these contexts. Note that *whose* is twice as frequent in these formulaic phrases than anywhere else, but it is still rare. The phrase *the felony for w'ch she stood indicted in* was repeated twice, selecting ‘which’ both times, and interestingly contains an instance of mismatched preposition doubling (see Riley & Parker, 1986). The wording is consistent enough to suggest that this phrasing is specific to legal language and contexts. The set phrase *to all whom it may concerne* appears in letters to multiple people and in petitions, and categorically selects ‘whom’. Inclusion of these tokens would artificially increase the rates of ‘whom’.

There were also several references to, and quotes from, scripture. Given that these were repetitions of phrases that speakers would have either read or heard regularly, they were excluded on account of their being unrepresentative of the target vernacular. There were two tokens, one with *that* and one with *which*.

Two of the set phrases demonstrated variability that might motivate their inclusion in the data. References to *a person REL tould me* were frequent in witness depositions. The wording was often nearly identical, with some variation in the choice of relative marker. *That* was chosen 61% of the time, while *which* was found in 39% of cases. Zero is not found in this context, but as a subject relative this might accord with claims in the literature that zero is disfavoured in subject relatives (Ball, 1996: 247; Johansson, 2017: 274). The set phrase *the oath REL they had taken* favours selection of *which* (54%). *That* is disfavoured, only being selected in 8% of cases. Zero occupies a middle ground, being selected 38% of the time.

There appears to be a clear division among the set phrases – there are those that do not display any variability, and are fully conventionalised (e.g. *we whose names are underwritten, all that is above written*), and there are those which continue to display variability – *a person REL tould me* shows variation between *which* and *that*, while the phrase *the oath REL they had taken* shows variability between *which*, *that*, and zero. These set phrases do not yet appear to be fully conventionalised, motivating their (re)inclusion in the data.

### 3.4.2 Syntactic Function of the Relative Marker

It is well established in the literature that subject relatives pattern differently from non-subject relatives in terms of relativiser selection (Ball, 1996). In particular, in contemporary vernaculars, zero subject relatives are rare and primarily restricted to *there*-existentials and *it*-clefts (Ball, 1996; D’Arcy & Tagliamonte, 2015; Tagliamonte, 2013; Tagliamonte et al., 2005). The remainder of the analysis will therefore consider subject and non-subject relative constructions separately. Table 4 below displays the distribution of variants by syntactic function of the relativiser.

| Factors        | <i>who</i> |    | <i>whom</i> |   | <i>whose</i> |   | <i>which</i> |    | <i>that</i> |     | zero |    | other |   | <b>Total</b> |            |
|----------------|------------|----|-------------|---|--------------|---|--------------|----|-------------|-----|------|----|-------|---|--------------|------------|
|                | %          | N  | %           | N | %            | N | %            | N  | %           | N   | %    | N  | %     | N | %            | N          |
| <i>Subject</i> | 7          | 31 | 0           | 0 | 0            | 0 | 13           | 59 | 73          | 322 | 6    | 25 | 1     | 4 | <b>59</b>    | <b>441</b> |

|                 |          |           |          |          |          |          |           |            |           |            |           |            |          |          |            |            |
|-----------------|----------|-----------|----------|----------|----------|----------|-----------|------------|-----------|------------|-----------|------------|----------|----------|------------|------------|
| <i>Object</i>   | 0        | 0         | 1        | 2        | 0        | 0        | 25        | 51         | 31        | 63         | 42        | 86         | 1        | 2        | <b>27</b>  | <b>204</b> |
| <i>Oblique</i>  | 0        | 0         | 2        | 2        | 0        | 0        | 30        | 28         | 27        | 25         | 39        | 36         | 2        | 2        | <b>13</b>  | <b>93</b>  |
| <i>Genitive</i> | 0        | 0         | 0        | 0        | 100      | 5        | 0         | 0          | 0         | 0          | 0         | 0          | 0        | 0        | <b>1</b>   | <b>5</b>   |
| <b>Total</b>    | <b>4</b> | <b>31</b> | <b>1</b> | <b>4</b> | <b>1</b> | <b>5</b> | <b>19</b> | <b>138</b> | <b>55</b> | <b>410</b> | <b>20</b> | <b>147</b> | <b>1</b> | <b>8</b> | <b>100</b> | <b>743</b> |

Table 4: Contribution of syntactic function of the relative marker to the selection of variants

As can be seen, breaking down the data by syntactic function reveals some key differences. Recall that for overall distribution, zero is responsible for 20% of tokens. In Table 4, zero is shown to be disfavoured in subject position with a rate of 6%. Rates of zero in non-subject positions (average of 40%) are twice as high as the overall distribution (20%). Interpreting rates relative to the overall distribution allows for the identification of which factors favour or disfavour any given variant. If a factor results in a higher rate of that particular variant than the overall distribution, then this factor may be said to favour this variant (or disfavour, if the rate is lower).

*Who* is entirely restricted to subject position in these data. *Whom* and *whose*, though much less frequent, are the non-subject and genitive counterparts respectively. These forms are all relatively rare – *who* is more common than *whom* and *whose* by a notable margin, but still only accounts for 7% of subject relatives in the data. Genitive constructions are very infrequent, but categorically select the relative marker *whose*. Due to their relative scarcity in the data, genitives will not be discussed further.

*Which* is favoured in non-subject relatives, where it is roughly on par with rates for *that* (25-30% in both object and oblique relatives). It is disfavoured in subject position at 13%. This underscores the importance of analysing subject and non-subject relatives separately. A study that only looked at subject relatives (e.g. Ball, 1996) might unintentionally predict lower rates for *which* than is in fact the case. This effect would be obscured if subject and non-subject relatives were not examined separately.

*That* is the dominant relativiser in subject position, occurring 73% of the time. *That* is disfavoured in non-subject positions, with rates of 31% in object relatives and 25% in obliques. Zero is in direct contrast to this, where it is the relative marker of choice in non-subject positions (42% in object relatives, 39% in obliques). This follows claims made in the literature (e.g. Tottie & Johansson, 2015) that by the 17<sup>th</sup> century the zero relative had shifted away from the subject position, becoming confined primarily to non-subject relatives. Nonetheless, zero subject relatives are clearly present in the data. Previous claims (e.g. Ball, 1996; D’Arcy & Tagliamonte, 2015; Tagliamonte, 2013; Tagliamonte et al., 2005; Tottie & Johansson, 2015) about zero subject relatives have associated their appearance in Early Modern English with specific constructions (*there*-existentials and *it*-clefts). As will be shown, this generalisation also holds true in the data from the *Salem Witchcraft Papers*.

Having outlined the effect of grammatical role of the relativiser on the selection of variant markers, it will now be possible to evaluate the influence of several factor groups that have been connected with variant selection, and the interplay between those factor groups and the grammatical role of the relative.

### 3.4.3 Subject Relativisers

#### 3.4.3.1 Animacy

A key factor group that is considered to exert a strong influence over the selection of relative markers is the animacy of the antecedent. This is particularly relevant to the choice between *wh*-relatives (*who* for human antecedents, *which* for non-human), but there are also claims (e.g. Ball, 1996) that relativiser *that* becomes more associated with non-human antecedents in this period. As Table 5 below demonstrates, the animacy of the antecedent head NP has some effect on the selection of variants in subject relatives.

| Factors          | <i>who</i> |    | <i>which</i> |    | <i>that</i> |     | zero |    | other |   | <b>Total</b> |     |
|------------------|------------|----|--------------|----|-------------|-----|------|----|-------|---|--------------|-----|
|                  | %          | N  | %            | N  | %           | N   | %    | N  | %     | N | %            | N   |
| <i>human</i>     | 11         | 31 | 9            | 26 | 74          | 219 | 5    | 15 | 1     | 3 | 67           | 294 |
| <i>non-human</i> | 0          | 0  | 22           | 33 | 70          | 103 | 7    | 10 | 1     | 1 | 33           | 147 |
| <b>Total</b>     | 7          | 31 | 13           | 59 | 73          | 322 | 6    | 25 | 1     | 4 | 100          | 441 |

Table 5: Contribution of animacy of the antecedent to the selection of variants in subject relatives



*That* dominates each sub-division within the table, and appears to be animacy neutral in subject position with rates around 70% (but is slightly more common with human antecedents at 74%), suggesting that the *that* relative marker was not yet predominantly confined to marking non-human antecedents (Ball, 1996). Similarly, zero shows little differentiation in terms of animacy, with the caveat that there are low rates of zero in subject position.

*Who* is exclusively found with human antecedents. However, even in that context, the rate of *who* is low (11%), and is almost on par with *which* (9%). *Which* does not yet appear to be confined to non-human antecedents, as is the case in modern mainstream varieties of English (e.g. D’Arcy & Tagliamonte, 2010). *Which* in subject position is favoured with non-human antecedents (22%), and disfavoured with human antecedents (9%). In the case of subject *which*, then, this relative marker exhibits a propensity to correlate more strongly with non-human than human antecedents, but is still available for both.

These results pattern similarly to those reported by Ball (1996: 247), who only discusses subject relatives with human antecedents in the *Salem Witchcraft Papers*. *That* is similarly found to dominate at 77%, followed by *which* at 11%, zero at 8%, and *who* at 4%. Beyond minor deviations of 2-3%, the main difference between the results here and those from Ball (1996) lies in the distribution of personal subject *who* (4% compared to my 11%). Ball (1996) reports 142 tokens of personal subject relatives – considerably fewer than my 294 tokens. I would suggest that this reflects changes in the availability of the target data. The current version of the *Salem Witchcraft Papers* includes more recently discovered documents that would not have been available in 1996. Another possibility might be that my inclusion of clefts (which Ball excluded) is inflating rates of *who* in my data. A closer examination reveals that this cannot be the case. Of the 294 personal subject relatives, only 40 tokens are clefts, and of these, none select *who*.

It seems clear that in the speech-related data in the *Salem Witchcraft Papers*, the animacy constraint proposed by Ball (1996) was not yet operative in subject relatives. *That* shows no marked animacy effects, while *which* is still present with human antecedents, at rates almost on par with *who*. This tallies with Ball’s (1996: 247-248) claim that the animacy constraint is predominantly literary at this point in time,

and had not yet infiltrated speech-like texts. Ball (1996: 248) does note that in the *Salem Witchcraft Papers*, personal subject *which* almost exclusively appears in the formulaic phrase *a person which tould me*. This is not true in the data analysed here, where this construction only accounts for 44% of personal subject *which*.

### 3.4.3.2 Matrix Clause Type

Recall that zero subject relatives have been connected with very specific matrix-clause constructions in the literature – namely, *there*-existentials and clefts (Tottie & Johansson, 2015). In order to test this, all tokens were coded for matrix clause type. As can be seen, a majority of tokens belong to the ‘other’ category, and so existentials, clefts, and stative possessives represent a relative minority of constructions. Existentials and clefts in particular did reveal some noteworthy patterns.

| Factors                   | <i>who</i> |           | <i>which</i> |           | <i>that</i> |            | zero     |           | other    |          | <b>Total</b> |            |
|---------------------------|------------|-----------|--------------|-----------|-------------|------------|----------|-----------|----------|----------|--------------|------------|
|                           | %          | N         | %            | N         | %           | N          | %        | N         | %        | N        | %            | N          |
| <i>Existential</i>        | 0          | 0         | 12           | 3         | 60          | 15         | 28       | 7         | 0        | 0        | 6            | 25         |
| <i>Cleft</i>              | 0          | 0         | 0            | 0         | 85          | 41         | 15       | 7         | 0        | 0        | 11           | 48         |
| <i>Stative Possessive</i> | 7          | 2         | 14           | 4         | 75          | 21         | 0        | 0         | 4        | 1        | 6            | 28         |
| <i>Other</i>              | 9          | 29        | 15           | 52        | 72          | 245        | 3        | 11        | 1        | 3        | 77           | 340        |
| <b>Total</b>              | <b>7</b>   | <b>31</b> | <b>13</b>    | <b>59</b> | <b>73</b>   | <b>322</b> | <b>6</b> | <b>25</b> | <b>1</b> | <b>4</b> | <b>100</b>   | <b>441</b> |

Table 6: Contribution of matrix clause construction type to the selection of variants in subject relatives

The *wh*-pronouns seem unaffected by matrix clause type in subject relatives. *Which* does not appear in cleft constructions, which may just reflect very low token counts for these constructions rather than a strong constraint. Relative *that* represents a mixed picture. It is slightly favoured with clefts (85%), and disfavoured with existentials (60%), though it still dominates across the board. Rates for *that* with stative possessives mirror overall rates for *that* in subject position. The zero relative is strongly favoured with existentials (28%), closely followed by *it*-clefts (15%), and is entirely absent from stative possessives.

These findings are broadly aligned with claims in the literature that the zero subject relative is more closely associated with existentials and clefts. Drawing on trial data from the *Corpus of English Dialogues* 1560-

1760, Tottie and Johansson (2015) found that in existential constructions, zero was selected 20% of the time, somewhat less than my rate of 28%. Clefts were much less common in their data, and so the distribution of variants in clefts was not examined. Given the results presented here, it might be reasonable to propose that the effect of matrix clause type on relative selection is stronger with existentials than with clefts (though both favour the selection of zero).

#### 3.4.3.3 Processing Constraints

Processing constraints have been proposed to exert a strong influence on relativiser selection (e.g. Quirk, 1957; Rohdenburg, 1996). In order to test this, two factor groups were operationalized – length of the relative clause (with longer clauses equating with more complexity) and adjacency. To calculate clause length, the number of orthographic words in each relative clause was counted (discounting the relative marker, if present). The clauses fell neatly into three broad patterns – short (1-5 words), medium (6-9 words), and long (10+ words).

| Factors       | <i>who</i> |           | <i>which</i> |           | <i>that</i> |            | zero     |           | other    |          | <b>Total</b> |            |
|---------------|------------|-----------|--------------|-----------|-------------|------------|----------|-----------|----------|----------|--------------|------------|
|               | %          | N         | %            | N         | %           | N          | %        | N         | %        | N        | %            | N          |
| <i>Short</i>  | 6          | 17        | 9            | 24        | 77          | 216        | 8        | 23        | 1        | 2        | <b>64</b>    | <b>282</b> |
| <i>Medium</i> | 8          | 9         | 14           | 16        | 76          | 88         | 2        | 2         | 1        | 1        | <b>26</b>    | <b>116</b> |
| <i>Long</i>   | 12         | 5         | 43           | 18        | 43          | 18         | 0        | 0         | 2        | 1        | <b>10</b>    | <b>42</b>  |
| <b>Total</b>  | <b>7</b>   | <b>31</b> | <b>13</b>    | <b>59</b> | <b>73</b>   | <b>322</b> | <b>6</b> | <b>25</b> | <b>1</b> | <b>4</b> | <b>100</b>   | <b>441</b> |

Table 7: Contribution of length of the relative clause to the selection of variants in subject relatives

In subject relatives, long clauses favoured the use of the *wh*-pronouns *who* (12%) and *which* (43%). *Which* was disfavoured in short clauses (9%), while *that* was disfavoured in long clauses (43%). Zero was disfavoured in medium clauses (2%) and was entirely absent from long clauses (although with only 42 tokens, long subject clauses are considerably rarer in the data than short and medium length subject relatives). This finding accords with similar results found by Quirk (1957: 108) for present day educated spoken British English, where zero subject relatives are confined to much shorter clauses (2-3 words), with almost no clauses longer than 6 words. *That* shows usage up to 9 words, but little above this, while *which* is prevalent in longer clauses.

Another processing constraint that has been connected with relative selection is adjacency of the relative marker to its antecedent. The zero relative in particular is claimed to be favoured in adjacent contexts. In the data from the *Salem Witchcraft Papers*, in both subject and non-subject constructions, immediate adjacency is by far the most common construction, accounting for 88% of all tokens.

| Factors             | <i>who</i> |    | <i>which</i> |    | <i>that</i> |     | zero |    | other |   | <b>Total</b> |     |
|---------------------|------------|----|--------------|----|-------------|-----|------|----|-------|---|--------------|-----|
|                     | %          | N  | %            | N  | %           | N   | %    | N  | %     | N | %            | N   |
| <i>Adjacent</i>     | 6          | 25 | 13           | 49 | 74          | 289 | 6    | 25 | 1     | 3 | 88           | 391 |
| <i>Non-adjacent</i> | 12         | 6  | 20           | 10 | 66          | 33  | 0    | 0  | 2     | 1 | 11           | 50  |
| <b>Total</b>        | 7          | 31 | 13           | 59 | 73          | 322 | 6    | 25 | 1     | 4 | 100          | 441 |

Table 8: Contribution of adjacency to the selection of variants in subject relatives

In subject relatives, *that* is the relative marker of choice regardless of adjacency. It is slightly disfavoured in non-adjacent contexts (66%), but when adjacent, rates are on par with overall distribution (74%). *Which* is favoured with non-adjacent contexts (20%). *Who* is favoured in non-adjacent (12%) contexts, while zero is exclusively found in adjacent contexts. Tottie and Johansson (2015) found similar results, reporting rates of 40% for zero subject relatives in adjacent contexts, and 22% in near-adjacent contexts (1-4 words).

<sup>11</sup> When the number of words is greater than 4, subject zero is entirely absent from the data. Quirk (1957: 107), looking at spoken educated Present Day English, found a similar pattern, where *which* is predominant in non-adjacent relative clauses, *that* is disfavoured, and zero is almost non-existent.

The above data suggest that processing constraints play a notable role in relativiser selection. When looking at the length of the relative clause, there is a divide between the *wh*-relatives and the traditional relatives (*that* and zero), with the former appearing more frequently in longer clauses. *That* and zero on the other hand are notably restricted to shorter clauses, with this effect being particularly strong for zero relatives. *That* is equally frequent with short and medium length clauses, but zero is disfavoured with medium length. Adjacency of the relative clause and the antecedent demonstrates the same division – *wh*-

<sup>11</sup> Tottie & Johansson (2015) include data from both trials and comedy. Their much higher rates for zero subject relatives are attributable to the generally higher rates of these constructions in comedy (50% compared to 20% in trials).

relatives are favoured in non-adjacent contexts, while *that* and zero are disfavoured when non-adjacent. Again, the effect is stronger for zero than it is for *that*, which is only slightly disfavoured in non-adjacent contexts.

The question remains how to explain these facts. The consistency of these results with those of Tottie and Johansson (2015) and particularly Quirk (1957) suggest a strong internal constraint. It is striking that the same patterns that obtain in Early Modern English are also present in (educated) Present Day English. Rohdenburg's (1996: 151) complexity principle, reproduced below, may go some way towards explaining these facts.

“In the case of more or less explicit grammatical options the more explicit one(s) will tend to be favored in cognitively more complex environments.”

Rohdenburg (1996) connects the use of the zero relative with less cognitively complex environments, i.e. adjacent contexts. Longer relative clauses are considered to increase cognitive complexity, raising processing demands and therefore favouring more explicit relative markers. This quite neatly dovetails with the results found here: the *wh*-relatives are favoured in the most complex positions (non-adjacency, longer relative clauses), while the zero relative is favoured in the least complex (adjacent, short relative clauses). *That* occupies a middle ground. It is an explicit relative, but without the complexity associated with *wh*-relatives. Therefore it follows that *that* would be only slightly disfavoured in non-adjacent contexts, and would still show high rates with medium length relative clauses. Based on the results for both the length of the relative clause and adjacency of the relative marker and antecedent, processing costs seem to play a clear role in relativiser selection, disproportionately affecting the zero relative marker.

#### 3.4.3.4 Antecedent Type

Another factor group that was investigated was the type of antecedent – whether pronominal, or a noun phrase. Table 9 shows the results for this factor group with subject relatives.

| Factors            | <i>who</i> |    | <i>which</i> |    | <i>that</i> |     | zero |    | other |   | <b>Total</b> |            |
|--------------------|------------|----|--------------|----|-------------|-----|------|----|-------|---|--------------|------------|
|                    | %          | N  | %            | N  | %           | N   | %    | N  | %     | N | %            | N          |
| <i>noun phrase</i> | 8          | 25 | 15           | 51 | 72          | 236 | 5    | 16 | 1     | 2 | <b>75</b>    | <b>330</b> |

|                |          |           |           |           |           |            |          |           |          |          |            |            |
|----------------|----------|-----------|-----------|-----------|-----------|------------|----------|-----------|----------|----------|------------|------------|
| <i>pronoun</i> | 5        | 6         | 7         | 8         | 77        | 86         | 8        | 9         | 2        | 2        | <b>25</b>  | <b>111</b> |
| <b>Total</b>   | <b>7</b> | <b>31</b> | <b>13</b> | <b>59</b> | <b>73</b> | <b>322</b> | <b>6</b> | <b>25</b> | <b>1</b> | <b>4</b> | <b>100</b> | <b>441</b> |

Table 9: Contribution of antecedent type to the selection of variants in subject relatives

*That* is slightly favoured with pronominal antecedents, while *who* is slightly disfavoured with pronominal antecedents. *Which* is disfavoured with pronominal antecedents (7%), and is twice as likely to appear with a nominal antecedent. Quirk (1957: 108) claims that zero subject relatives are more likely to have a pronominal antecedent. This claim is borne out in my data, where zero is slightly favoured with pronominal antecedents (8%).

The effect of definiteness on the selection of variants was also tested. In subject relatives, zero shows no effect. The *wh*-relatives pattern similarly – both *who* and *which* are favoured with indefinite antecedents, and disfavoured with definite antecedents, though this effect is slight. The most notable result for subject relatives is the slight preference for *that* with definite antecedents (81%), and its disfavouring with indefinite antecedents (62%).

| Factors           | <i>who</i> |           | <i>which</i> |           | <i>that</i> |            | zero     |           | other    |          | <b>Total</b> |            |
|-------------------|------------|-----------|--------------|-----------|-------------|------------|----------|-----------|----------|----------|--------------|------------|
|                   | %          | N         | %            | N         | %           | N          | %        | N         | %        | N        | %            | N          |
| <i>Definite</i>   | 4          | 10        | 10           | 24        | 81          | 204        | 5        | 13        | 0        | 1        | <b>57</b>    | <b>252</b> |
| <i>Indefinite</i> | 11         | 21        | 19           | 35        | 62          | 118        | 6        | 12        | 2        | 3        | <b>43</b>    | <b>189</b> |
| <b>Total</b>      | <b>7</b>   | <b>31</b> | <b>13</b>    | <b>59</b> | <b>73</b>   | <b>322</b> | <b>6</b> | <b>25</b> | <b>1</b> | <b>4</b> | <b>100</b>   | <b>441</b> |

Table 10: Contribution of definiteness of the antecedent to the selection of variants in subject relatives

Definiteness and antecedent type have received attention in the literature, with somewhat inconsistent results. For example, the zero relative has often been connected with pronominal antecedents (Huddleston, 1971; Piercy et al., 2011; Tottie, 1995; Tottie & Rey, 1997). Others have connected zero with indefinite noun phrase antecedents (Tagliamonte et al., 2005; Tottie & Harvie, 2000). Wasow et al. (2011) find zero to be more frequent with definite antecedents than indefinites, while Harvie (1998) finds that pronouns and indefinite noun phrases show little difference in the selection of zero. There is a possibility that this remarkable inconsistency (across varieties and time periods) lends some credence to Ball's (1996) claim

that there is no vernacular norm in the relative system. Tagliamonte et al. (2005) propose that the lack of consistency is a result of interaction between definiteness and matrix clause type, as a majority of existential constructions have indefinite antecedents. The differing rates reported in the literature are therefore claimed by Tagliamonte et al. (2005) to be a reflection of the proportion of existentials in the data (if they are not excluded). My data upholds this claim. While existentials are rare with only 30 tokens across subject and non-subject relatives, they take indefinite antecedents 83% of the time (25/30 tokens).

My data shows a slight preference for *wh*-relatives with indefinite and nominal antecedents, and for *that* with definite and pronominal antecedents. Zero is favoured with pronominal antecedents, but shows only minor preference for indefinite antecedents. Recall that the accessibility scale discussed in Section 3.3.2.5 (Ariel, 1990; Jaeger and Wasow, 2005) predicts that the zero relative is more likely with more accessible antecedents (e.g. pronouns). This tallies with claims by Rohdenburg (1996: 174) that pronouns are easier to process than definite NPs. The application of this accessibility scale to my results is equally instructive. The zero relative is indeed favoured with pronominal antecedents. The almost identical rates of zero for definites and indefinites could be related to two facts – low token counts, and their adjacency on the accessibility scale.

#### 3.4.3.5 Sex

Consideration will now be given to several social factor groups, to evaluate their influence on the selection of relative markers. For social factors, due to low rates of *who* and *whom* in particular, the *wh*-relatives have been combined under a single WH grouping. In cases where a speaker's social characteristics were missing from the historical record or could not be reliably established, tokens were coded as *unknown*.

| Factors        | WH        |           | <i>that</i> |            | zero     |           | other    |          | Total      |            |
|----------------|-----------|-----------|-------------|------------|----------|-----------|----------|----------|------------|------------|
|                | %         | N         | %           | N          | %        | N         | %        | N        | %          | N          |
| <i>male</i>    | 22        | 46        | 70          | 146        | 6        | 13        | 2        | 4        | 47         | 209        |
| <i>female</i>  | 19        | 37        | 74          | 143        | 6        | 12        | 0        | 0        | 44         | 192        |
| <i>unknown</i> | 18        | 7         | 83          | 33         | 0        | 0         | 0        | 0        | 9          | 40         |
| <b>Total</b>   | <b>20</b> | <b>90</b> | <b>73</b>   | <b>322</b> | <b>6</b> | <b>25</b> | <b>1</b> | <b>4</b> | <b>100</b> | <b>441</b> |

Table 11: Contribution of speaker sex to the selection of variants in subject relatives

Subject relatives demonstrate only slight differences in the distribution of variants between male (who slightly favour *wh*-relatives) and female speakers (who slightly favour *that*). For each variant, the difference is under 5%, and therefore sex does not appear to be a strong determinant of relative marker selection. These results stand in stark contrast to Huber (2017), who finds that sex is a significant factor in relativiser selection. In Huber's (2017) British trial data from the Old Bailey Corpus, female speakers consistently favour the use of zero across time periods (rates of 29% in the first period, 1720-1789). Male speakers are found to favour the use of *wh*-relatives (29% in the first period). It should be noted that Huber (2017) does not separate subject and non-subject relatives when discussing sex, and so there is the possibility that non-subject relatives in my data will show sex effects. I return to this issue below.

#### 3.4.3.6 Birthplace

Analysis of the effects of speaker birthplace on relativiser selection is hampered somewhat by low token counts for speakers born outside of New England. Additionally, for 33% of subject relatives, the birthplace of the speaker is either not known or could not be reliably determined.

| Factors              | WH        |           | <i>that</i> |            | zero     |           | other    |          | Total      |            |
|----------------------|-----------|-----------|-------------|------------|----------|-----------|----------|----------|------------|------------|
|                      | %         | N         | %           | N          | %        | N         | %        | N        | %          | N          |
| <i>South America</i> | 17        | 2         | 67          | 8          | 8        | 1         | 0        | 0        | 3          | 12         |
| <i>New England</i>   | 23        | 56        | 70          | 168        | 5        | 12        | 2        | 4        | 54         | 240        |
| <i>Britain</i>       | 12        | 5         | 81          | 35         | 7        | 3         | 0        | 0        | 10         | 43         |
| <i>Unknown</i>       | 18        | 27        | 76          | 111        | 6        | 9         | 0        | 0        | 33         | 146        |
| <b>Total</b>         | <b>20</b> | <b>90</b> | <b>73</b>   | <b>322</b> | <b>6</b> | <b>25</b> | <b>1</b> | <b>4</b> | <b>100</b> | <b>441</b> |

Table 12: Contribution of speaker birthplace to the selection of variants in subject relatives

The data contains a single speaker from South America (Tituba). It would therefore not be possible to claim that these results are representative. It is however interesting to note that rates are in line with overall rates for subject relatives across variants. The most notable difference here is between the use of *that* and WH among speakers from New England and those born in Britain. Use of *that* is favoured among speakers



born in Britain (81%), while WH is slightly disfavoured (12%). It should be noted that with only 43 tokens from these speakers, it is not clear that this defines a transatlantic difference in relativiser selection. However, comparison with 17<sup>th</sup> century British English data (Levey, 2017) in the *Corpus of English Dialogues* (1560-1760) supports this finding, as Levey (2017) also reports a rate of 12% for *wh*-relatives in subject position.

### 3.4.3.7 Age

The data were also tested for the effect of speaker age on the selection of variants. The logic here is that if the relative system is undergoing change in 17<sup>th</sup> century Massachusetts English, then this might be reflected in the language of younger speakers.

| Factors      | WH        |           | <i>that</i> |            | zero     |           | other    |          | Total      |            |
|--------------|-----------|-----------|-------------|------------|----------|-----------|----------|----------|------------|------------|
|              | %         | N         | %           | N          | %        | N         | %        | N        | %          | N          |
| Younger      | 31        | 33        | 67          | 71         | 2        | 2         | 0        | 0        | 24         | 106        |
| Middle       | 24        | 32        | 69          | 93         | 4        | 5         | 3        | 4        | 30         | 134        |
| Older        | 11        | 12        | 79          | 83         | 10       | 10        | 0        | 0        | 24         | 105        |
| Unknown      | 14        | 13        | 78          | 75         | 8        | 8         | 0        | 0        | 22         | 96         |
| <b>Total</b> | <b>20</b> | <b>90</b> | <b>73</b>   | <b>322</b> | <b>6</b> | <b>25</b> | <b>1</b> | <b>4</b> | <b>100</b> | <b>441</b> |

Table 13: Contribution of speaker age to the selection of variants in subject relatives

The data is suggestive of a trend whereby the *wh*-relatives are increasing at the expense of *that* and zero. *That* continues to dominate across age groups, but with a rate of 79%, it is most frequent among the eldest speakers (50+). This rate steadily decreases as speakers get younger – 69% for 30-49, and 67% for 0-29, both slightly below the overall distribution for *that*. A similar situation obtains for zero subject relatives – they are most commonly used by older speakers (10%), followed by the middle group (4%) and finally are used least by the youngest group (2%). In contrast, WH is disfavoured among the eldest speakers (11%). Rates are in line with the overall distribution for speakers aged 30-49 (24), while younger speakers use WH 31% of the time.

Table 14 below shows how this effect is distributed across the *wh*-relatives. *Which* adheres to the pattern, being favoured among younger speakers (24%) and disfavoured by older speakers (9%). The effect is less clear for *who*: while older speakers disfavour the use of *who* (3%), the greatest frequency of *who* is with the middle age group (13%). The youngest speakers use *who* 8% of the time.

| Factors        | <i>who</i> |    | <i>which</i> |    | <i>that</i> |     | zero |    | other |   | <b>Total</b> |     |
|----------------|------------|----|--------------|----|-------------|-----|------|----|-------|---|--------------|-----|
|                | %          | N  | %            | N  | %           | N   | %    | N  | %     | N | %            | N   |
| <i>Young</i>   | 8          | 8  | 24           | 25 | 67          | 71  | 2    | 2  | 0     | 0 | 24           | 106 |
| <i>Middle</i>  | 10         | 13 | 14           | 19 | 69          | 93  | 4    | 5  | 3     | 4 | 30           | 134 |
| <i>Old</i>     | 3          | 3  | 9            | 9  | 79          | 83  | 10   | 10 | 0     | 0 | 24           | 105 |
| <i>Unknown</i> | 7          | 7  | 6            | 6  | 78          | 75  | 8    | 8  | 0     | 0 | 22           | 96  |
| <b>Total</b>   | 7          | 31 | 13           | 59 | 73          | 322 | 6    | 25 | 1     | 4 | 100          | 441 |

Table 14: Contribution of speaker age to the selection of variants in subject relatives (*who* and *which* separated)

Huber (2017) finds that in subject relatives, the *wh*-relatives are increasing at the expense of *that*. Bearing in mind that Huber's (2017) data is from the Old Bailey Corpus and therefore reflects British English usage, it is interesting to note that my data suggests that it is only *which* that is increasing, as *who* shows the greatest use among the middle-aged group of speakers. There is no evidence of a monotonic trend whereby *who* increases as the age spectrum is descended. However, with only 31 tokens of *who*, I exercise caution in making strong claims about changes in the use of *who*.

Tottie and Johansson (2015: 140), looking at trial data from the *Corpus of English Dialogues* 1560-1760, also find a decrease in rates of zero subject relatives after period 3 (1640-1679), which coincides with the period that the older and middle-aged speakers in the *Salem Witchcraft Papers* would have been growing up. While it is not possible to draw any definitive conclusions on the basis of only 25 tokens of the zero subject relativiser, it is nonetheless interesting to note a possible increase in *which* at the expense of *that* and zero. Further investigation with a larger amount of data would be required to verify if this is indeed the case.

#### 3.4.3.8 Social Status

The *wh*-relatives have been proposed to have been a change from above (e.g. Romaine, 1982), originating in the upper echelons of society. If this is indeed the case, then it might be expected that in 17<sup>th</sup> century Massachusetts, some 200 years after the first appearance of relative *who*, the *wh*-relatives will show greater distribution among speakers belonging to Group A (the highest status speakers), while the more colloquial zero subject relative will be more frequent among speakers from Groups B and C (see Section 3.3.3.4 for a discussion on how speakers were assigned to each group). Group D contains a very small number of tokens, from only two speakers, and so will not be discussed further.

| Factors      | WH        |           | <i>that</i> |            | zero     |           | other    |          | Total      |            |
|--------------|-----------|-----------|-------------|------------|----------|-----------|----------|----------|------------|------------|
|              | %         | N         | %           | N          | %        | N         | %        | N        | %          | N          |
| Group A      | 12        | 9         | 77          | 56         | 7        | 5         | 4        | 3        | 17         | 73         |
| Group B      | 26        | 32        | 71          | 87         | 2        | 3         | 1        | 1        | 28         | 123        |
| Group C      | 26        | 15        | 66          | 38         | 9        | 5         | 0        | 0        | 13         | 58         |
| Group D      | 15        | 2         | 77          | 10         | 8        | 1         | 0        | 0        | 3          | 13         |
| Unknown      | 18        | 32        | 75          | 131        | 6        | 11        | 0        | 0        | 39         | 174        |
| <b>Total</b> | <b>20</b> | <b>90</b> | <b>73</b>   | <b>322</b> | <b>6</b> | <b>25</b> | <b>1</b> | <b>4</b> | <b>100</b> | <b>441</b> |

Table 15: Contribution of speaker social status to the selection of variants in subject relatives

Contrary to the hypothesised result, WH is disfavoured by Group A (12%), and slightly favoured by Groups B and C (26% each). Considering that the results from speaker age indicate an increase in *which* over time, it may be that speakers from Groups B and C (who favour the use of WH) are leading this change. This would align with modern sociolinguistic findings (e.g. Guy, 1989: 59; Guy et al., 1986; Labov, 1974, 1980, 1981) that the lower classes of the social hierarchy are responsible for innovating and advancing linguistic change.

*That* is fairly stable throughout, but is slightly disfavoured by Group C (66%). Group B disfavours the use of zero (2%). Given the low rates of zero subject relatives overall, this is not a particularly notable finding. Tottie and Johansson (2015), who used a binary distinction between upper class (nobility, clergymen, and judiciary) and everyone else, found no difference in rates of zero subject relatives. Differences in the

selection of variants between speaker groups is relatively minor. There is no clearly discernible effect in the selection of the zero relative marker in subject position. It is entirely likely that this is a result of low token counts for zero subject relatives (only 25 tokens).

It is also possible this data actually reflects the impact (and distribution) of speaker roles in the trials, rather than their place in the social hierarchy. Here, the context of the data is key: these are trials, with their own power dynamics, particularly between the judiciary officials, the accusers, and those who stand accused. Of note here is the fact that different ranks in the social hierarchy are unevenly distributed across trial roles: Group A predominates in the role of official (73% of tokens), while the speakers of Group C are primarily accusers (64%), and less frequently accused (24%). Group B show a fairly even distribution between accuser (31%) and witness (28%), but most frequently belong to the group of accused (39%).

Working on the assumption that this unique social context and dynamic might influence the language used, I investigated the distribution of variants by speaker role (Table 16). Speakers are coded as accused, accusers, officials, and witnesses (recorders of testimonials and writers of petitions). Here it can be seen that WH is slightly favoured among accusers (26%), and slightly disfavoured among officials (13%). *That* is stable across the board, and zero shows only minor variation. The differences here are relatively small, and suggest that speaker role is in fact not a determining factor of variant selection.

| Factors         | WH        |           | <i>that</i> |            | zero     |           | other    |          | Total      |            |
|-----------------|-----------|-----------|-------------|------------|----------|-----------|----------|----------|------------|------------|
|                 | %         | N         | %           | N          | %        | N         | %        | N        | %          | N          |
| <i>Accused</i>  | 19        | 23        | 73          | 87         | 8        | 9         | 0        | 0        | 27         | 119        |
| <i>Accuser</i>  | 26        | 22        | 69          | 59         | 4        | 3         | 1        | 1        | 19         | 85         |
| <i>Official</i> | 13        | 8         | 79          | 48         | 5        | 3         | 3        | 2        | 14         | 61         |
| <i>Witness</i>  | 24        | 28        | 73          | 86         | 3        | 3         | 1        | 1        | 27         | 118        |
| <i>Unknown</i>  | 16        | 9         | 72          | 42         | 12       | 7         | 0        | 0        | 13         | 58         |
| <b>Total</b>    | <b>20</b> | <b>90</b> | <b>73</b>   | <b>322</b> | <b>6</b> | <b>25</b> | <b>1</b> | <b>4</b> | <b>100</b> | <b>441</b> |

Table 16: Contribution of speaker role to the selection of variants in subject relatives

#### 3.4.3.9 Document Type

When examining the effect of document type, there is a clear continuum from the most speech-like texts (i.e. examinations) to the least speech-like (i.e. official documents). The *wh*-relatives are disfavoured in examinations (11%), and favoured in letters and official documents (35% and 60% respectively). The reverse is true for *that*, which is favoured in examinations (79%) and disfavoured in letters and official documents (60% and 40% respectively). Though rare in subject relatives, zero occurs most frequently in examinations (10%). With the exception of subject zero relatives, the rate for each variant in depositions patterns almost identically with the overall rate for each variant in subject relatives.

The distribution of variants across document types in subject relatives shows a clear preference for *wh*-relatives in less speech-like text types. This supports Romaine's (1982) claim that the *wh*-relatives have not made major inroads into the spoken language.

| Factors                   | WH        |           | <i>that</i> |            | zero     |           | other    |          | Total      |            |
|---------------------------|-----------|-----------|-------------|------------|----------|-----------|----------|----------|------------|------------|
|                           | %         | N         | %           | N          | %        | N         | %        | N        | %          | N          |
| <i>Examinations</i>       | 11        | 21        | 79          | 147        | 10       | 19        | 0        | 0        | 42         | 187        |
| <i>Depositions</i>        | 24        | 46        | 72          | 139        | 3        | 5         | 1        | 2        | 44         | 192        |
| <i>Letters</i>            | 35        | 20        | 60          | 34         | 2        | 1         | 4        | 2        | 13         | 57         |
| <i>Official Documents</i> | 60        | 3         | 40          | 2          | 0        | 0         | 0        | 0        | 1          | 5          |
| <b>Total</b>              | <b>20</b> | <b>90</b> | <b>73</b>   | <b>322</b> | <b>6</b> | <b>25</b> | <b>1</b> | <b>4</b> | <b>100</b> | <b>441</b> |

Table 17: Contribution of document type to the selection of variants in subject relatives

#### 3.4.4 Non-Subject Relativisers

##### 3.4.4.1 Animacy

Discussion will now turn to the non-subject relativisers (objects and obliques). The animacy constraint patterns in a notably different way for non-subject relatives. Results indicating the effect of this parameter on variant choice are given in Table 18 below.

| Factors | <i>whom</i> |   | <i>which</i> |   | <i>that</i> |   | zero |   | other |   | Total |   |
|---------|-------------|---|--------------|---|-------------|---|------|---|-------|---|-------|---|
|         | %           | N | %            | N | %           | N | %    | N | %     | N | %     | N |

|                  |          |          |           |           |           |           |           |            |          |          |            |            |
|------------------|----------|----------|-----------|-----------|-----------|-----------|-----------|------------|----------|----------|------------|------------|
| <i>human</i>     | 6        | 3        | 12        | 6         | 35        | 18        | 45        | 23         | 2        | 1        | 17         | 51         |
| <i>non-human</i> | 3        | 1        | 30        | 73        | 28        | 70        | 40        | 99         | 1        | 3        | 81         | 246        |
| <b>Total</b>     | <b>1</b> | <b>4</b> | <b>27</b> | <b>79</b> | <b>30</b> | <b>88</b> | <b>41</b> | <b>122</b> | <b>3</b> | <b>9</b> | <b>100</b> | <b>297</b> |

Table 18: Contribution of animacy of the antecedent to the selection of variants in non-subject relatives

*Who* is entirely absent in non-subject position in the data. Low rates prevent any definitive conclusions from being made about the distribution of *whom*, but it is primarily found with human antecedents (6%). There is a single token of *whom* with a non-human antecedent. Closer investigation reveals that in this token, the antecedent is an animal (61), perhaps suggesting a degree of personification.

- (61)        hee saw **a Great white dogg** whom he came up to  
(SWP063.18 – John Hughes,  
Testimony: March 5<sup>th</sup>, 1692)

*Which* is disfavoured (but still available) with human antecedents (12%), providing further evidence that the animacy constraint was only beginning to take shape in this variety at the end of the 17<sup>th</sup> century. *That* shows a very slight preference for human antecedents (35%) over non-human (28%). Similarly, rates for the zero relative are almost in line with its overall distribution in non-subject relatives. It seems clear that the animacy constraint posited by Ball (1996) was not yet fully operative in written approximations to spoken 17<sup>th</sup> century Massachusetts English.

These results contrast with those reported by Johansson (2017: 274), who finds rates of 41% for *whom* in personal non-subject relatives, compared to 2% *which*. This is a notable disparity, albeit one with a fairly simple explanation. Firstly, it is not clear that restrictive and non-restrictive relatives are treated separately in Johansson's (2017) data, which would artificially inflate rates of *wh*-forms. Similarly, Johansson's results conflate data from trials and dramas. It is entirely likely that these two genres pattern very differently, as Tottie and Johansson (2015) also conclude.

#### 3.4.4.2 Matrix Clause Type

In non-subject relatives, *whom* is only found in the ‘other’ category. However, extremely low rates for each clause type necessitate cautious interpretation of the data. *Which* is disfavoured with clefts and stative possessives (6% and 13% respectively), and is absent entirely from existentials. *That* is favoured in *there*-existentials (50%) and *it*-clefts (53%). The zero relative is disfavoured with existentials (25%), and slightly favoured with clefts (44%) and stative possessives (50%). This represents a direct contrast with subject zero relatives, which are favoured primarily with existentials (and to a lesser extent with clefts).

| Factors                    | <i>whom</i> |          | <i>Which</i> |           | <i>that</i> |           | zero      |            | other    |          | <b>Total</b> |            |
|----------------------------|-------------|----------|--------------|-----------|-------------|-----------|-----------|------------|----------|----------|--------------|------------|
|                            | %           | N        | %            | N         | %           | N         | %         | N          | %        | N        | %            | N          |
| <i>Existentials</i>        | 0           | 0        | 0            | 0         | 50          | 2         | 25        | 1          | 25       | 1        | 2            | 4          |
| <i>Clefts</i>              | 0           | 0        | 6            | 1         | 53          | 9         | 44        | 7          | 0        | 0        | 6            | 17         |
| <i>Stative Possessives</i> | 0           | 0        | 13           | 2         | 38          | 6         | 50        | 8          | 0        | 0        | 5            | 16         |
| <i>Other</i>               | 2           | 4        | 29           | 76        | 27          | 71        | 41        | 106        | 1        | 3        | 88           | 260        |
| <b>Total</b>               | <b>1</b>    | <b>4</b> | <b>27</b>    | <b>79</b> | <b>30</b>   | <b>88</b> | <b>41</b> | <b>122</b> | <b>1</b> | <b>4</b> | <b>100</b>   | <b>297</b> |

Table 19: Contribution of matrix clause construction type to the selection of variants in non-subject relatives

#### 3.4.4.3 Processing Constraints

The situation in non-subject relatives is much the same as with subject relatives. *Whom* is favoured increasingly as clauses become longer (3% in medium length, 6% in long clauses). *Which* shows a particular propensity to appear with long relative clauses (56%), and is slightly disfavoured in short (23%). *That* appears unaffected by clause length in non-subject relatives, and rates are in line with overall distribution regardless of clause length. Finally, zero is slightly disfavoured in medium length clauses (35%), and strongly disfavoured in long relative clauses (6%). As with subject relatives, long clauses are a minority, and so the results for these should be treated with caution.

| Factors | <i>whom</i> |   | <i>which</i> |   | <i>that</i> |   | zero |   | other |   | <b>Total</b> |   |
|---------|-------------|---|--------------|---|-------------|---|------|---|-------|---|--------------|---|
|         | %           | N | %            | N | %           | N | %    | N | %     | N | %            | N |

|               |          |          |           |           |           |           |           |            |          |          |            |            |
|---------------|----------|----------|-----------|-----------|-----------|-----------|-----------|------------|----------|----------|------------|------------|
| <i>Short</i>  | 0        | 1        | 23        | 49        | 30        | 64        | 46        | 99         | 1        | 3        | 73         | 216        |
| <i>Medium</i> | 3        | 2        | 32        | 20        | 29        | 18        | 35        | 22         | 2        | 1        | 21         | 63         |
| <i>Long</i>   | 6        | 1        | 56        | 10        | 33        | 6         | 6         | 1          | 0        | 0        | 6          | 18         |
| <b>Total</b>  | <b>1</b> | <b>4</b> | <b>27</b> | <b>79</b> | <b>30</b> | <b>88</b> | <b>41</b> | <b>122</b> | <b>1</b> | <b>4</b> | <b>100</b> | <b>297</b> |

Table 20: Contribution of length of the relative clause to the selection of variants in non-subject relatives

Turning now to adjacency, it can be seen that the processing constraints operative in subject relatives are also present with non-subject relatives. Non-adjacent contexts are notably less common than adjacent contexts.

| Factors             | <i>whom</i> |          | <i>which</i> |           | <i>that</i> |           | zero      |            | other    |          | <b>Total</b> |            |
|---------------------|-------------|----------|--------------|-----------|-------------|-----------|-----------|------------|----------|----------|--------------|------------|
|                     | %           | N        | %            | N         | %           | N         | %         | N          | %        | N        | %            | N          |
| <i>Adjacent</i>     | 1           | 2        | 21           | 56        | 31          | 80        | 46        | 120        | 2        | 4        | 88           | 262        |
| <i>Non-adjacent</i> | 6           | 2        | 66           | 23        | 23          | 8         | 6         | 2          | 0        | 0        | 12           | 35         |
| <b>Total</b>        | <b>1</b>    | <b>4</b> | <b>27</b>    | <b>79</b> | <b>30</b>   | <b>88</b> | <b>41</b> | <b>122</b> | <b>1</b> | <b>4</b> | <b>100</b>   | <b>297</b> |

Table 21: Contribution of adjacency to the selection of variants in non-subject relatives

*Whom*, though rare, appears to be favoured in non-adjacent contexts (6%). *Which* is strongly favoured in non-adjacent contexts (66%) and slightly disfavoured in adjacent contexts (21%). Zero is strongly disfavoured in non-adjacent contexts (6%). *That* is slightly disfavoured in non-adjacent contexts (23%) but rates in adjacent contexts are in line with overall distribution in non-subject relatives (31%).

Rohdenburg's (1996) Complexity Principle appears to apply to non-subject relatives in a similar fashion to subject relatives. It is interesting to note that the effect is notably stronger in non-subject relatives. The difference between rates of *which* in subject position (13% adjacent, 20% non-adjacent) versus non-subject position (21% adjacent and 66% non-adjacent) demonstrates this. Similarly, zero in subject position (6% adjacent and 0% non-adjacent) compared to non-subject position (46% adjacent and 6% non-adjacent) suggest that processing constraints have a stronger effect on non-subject relatives.



These findings align neatly with previous claims in the literature regarding an asymmetry in terms of processing constraints between subject and non-subject relative clauses (e.g. Gibson, 1998). In filler-gap dependencies (such as relative clauses), the distance between the moved element (the filler) and its trace (the gap) directly correlates with the complexity of the construction: greater distance results in greater complexity and increased processing costs (Hawkins, 1999). Under this logic, subject relative clauses have less distance between the filler and the gap; as a result, they are less complex (and therefore easier to process) than non-subject relatives.

#### 3.4.4.4 Preposition Stranding

It is instructive to consider in further detail the effect of preposition stranding/pied-piping on relativiser selection. This section necessarily focuses on a subset of the data, namely oblique relative clause constructions which admit variability between preposition stranding and pied-piping with *wh*-relativisers.

| Factors            | <i>whom</i> |          | <i>which</i> |           | <i>that</i> |           | zero      |           | other    |          | <b>Total</b> |           |
|--------------------|-------------|----------|--------------|-----------|-------------|-----------|-----------|-----------|----------|----------|--------------|-----------|
|                    | %           | N        | %            | N         | %           | N         | %         | N         | %        | N        | %            | N         |
| <i>Stranding</i>   | 1           | 1        | 15           | 11        | 34          | 25        | 49        | 36        | 0        | 0        | 78           | 73        |
| <i>Pied-piping</i> | 5           | 1        | 85           | 17        | 0           | 0         | 0         | 0         | 10       | 2        | 22           | 20        |
| <b>Total</b>       | <b>2</b>    | <b>2</b> | <b>30</b>    | <b>28</b> | <b>27</b>   | <b>25</b> | <b>39</b> | <b>36</b> | <b>2</b> | <b>2</b> | <b>100</b>   | <b>93</b> |

Table 22: Contribution of preposition stranding to the selection of variants

There is a clearly defined distinction in this data between *wh*-relatives, and *that* and zero. *That* and zero are entirely restricted to preposition stranding contexts, with the latter predominating (49% compared to 34% *that*). Yáñez-Bouza (2015) also reports that preposition stranding in Early Modern English is most frequent with a zero relative marker, but was still common with *that*. *Whom* shows a slight preference for pied-piping contexts (5%), but with only two tokens this is not a particularly conclusive finding. *Which* on the other hand is strongly favoured in pied-piping contexts (85%), and is by far the dominant relative in this context. *Which* is disfavoured in stranding contexts (15%), but it is still a viable option, refuting

the controversial claim by Van den Eynden (1996: 444) that stranding is not feasible with *wh*-relativisers in historical varieties of the language.

The overall rates for stranding reported here are much higher than those in the studies cited by Bergh and Seppänen (2000: 306), where stranding occurs around 11% of the time (compared to 89% pied-piping). The difference here lies in the type of data examined. Whereas much of the data from the *Salem Witchcraft Papers* affords a window on historical orality, the studies cited in Bergh and Seppänen (2000) more closely approximate conventional written genres. For example, Rydén (1966) looks at the works of several writers, including Sir Thomas Elyot; Ingels (1985) looks at essays and dramatic texts; Bengtsson (1996) draws on data from the works of Shakespeare; Lindelöf (1997) looks at the use of pied-piping and preposition stranding in the Early Modern English section of the Helsinki Corpus, which has a diverse range of text types including sermons, trial data, and educational treatises, among others. I propose that the considerably higher rates of pied-piping reported in these studies are a result of their proximity to the written standard, whereas the data from the *Salem Witchcraft Papers* examined here more closely approximates the spoken language of the 17<sup>th</sup> century.

Yáñez-Bouza (2015: 59), examining prescriptive writing in the 18<sup>th</sup> century, finds critical attitudes towards preposition stranding dating from as early as the 1710s, and notes that these attitudes would have possibly been familiar to speakers in the 17<sup>th</sup> century, as evidenced by the writings of John Dryden (1631-1700). It is therefore interesting to note, given the proscriptions against stranding the preposition which would find favour in the following decade, that preposition stranding was possibly the norm in 17<sup>th</sup> century Massachusetts English. In 78% of the oblique constructions in the data, the preposition was stranded. This raises the question whether these speakers were resistant to normative influence, or if stranding had always been the norm in speech-related sources, despite claims that it is not putatively feasible with *wh*-relatives (Van den Eynden, 1996).

Preposition stranding is subject to the influence of style and register, being favoured in informal, spoken language (Yáñez-Bouza, 2015: 152). This is important in the context of the *Salem Witch Trials*. Although the data are, to some extent, speech-based, they also represent a relatively formal register, in a court-room setting. There is every possibility then, that the actual rates for preposition stranding in informal spoken 17<sup>th</sup> century English were higher than those reported here.

#### 3.4.4.5 Antecedent Type

Table 23 shows the effect of grammatical category of the antecedent on relative marker choice in non-subject relatives. *Whom* and zero show only minor effects. Again, *which* is disfavoured (albeit only slightly) with pronominal antecedents (21%). Contrastively, *that* is slightly favoured with pronominal antecedents (36%). This effect is fairly weak, marking in both cases a 6% departure from the overall rate in non-subject relatives, and is not a strong determinant of variant choice.

| Factors            | <i>whom</i> |          | <i>which</i> |           | <i>that</i> |           | zero      |            | other    |          | <b>Total</b> |            |
|--------------------|-------------|----------|--------------|-----------|-------------|-----------|-----------|------------|----------|----------|--------------|------------|
|                    | %           | N        | %            | N         | %           | N         | %         | N          | %        | N        | %            | N          |
| <i>noun phrase</i> | 1           | 3        | 28           | 68        | 28          | 69        | 41        | 100        | 2        | 4        | <b>82</b>    | <b>244</b> |
| <i>pronoun</i>     | 2           | 1        | 21           | 11        | 36          | 19        | 42        | 22         | 0        | 0        | <b>18</b>    | <b>53</b>  |
| <b>Total</b>       | <b>1</b>    | <b>4</b> | <b>27</b>    | <b>79</b> | <b>30</b>   | <b>88</b> | <b>41</b> | <b>122</b> | <b>1</b> | <b>4</b> | <b>100</b>   | <b>297</b> |

Table 23: Contribution of grammatical category of the antecedent to the selection of variants in non-subject relatives

Similarly, definiteness has a weak effect on non-subject relativiser selection. *Whom* and *which* are essentially unaffected by definiteness. *That* is slightly favoured with indefinite antecedents (34%) and slightly disfavoured with definite antecedents (26%). The opposite is true for the zero relative, which is slightly favoured with definite antecedents (44%) and slightly disfavoured with indefinite antecedents (35%), but again these differences are at best marginal. Definiteness is therefore not a strong determinant of variant choice for non-subject relatives.

While much of the literature on the effects of antecedent type and definiteness is inconsistent regarding the patterns predicted, the received wisdom is that the zero relative will be preferred with pronominal and definite antecedents, due to lower associated processing costs (Rohdenburg, 1996: 174). My data adheres to this prediction, albeit fairly weakly. Zero is very slightly favoured with pronominal antecedents. The effect is particularly noticeable for definiteness, with zero being favoured with definite antecedents and disfavoured with indefinites. This effect is stronger for non-subject relatives than subject relatives.

| Factors           | <i>whom</i> |          | <i>which</i> |           | <i>that</i> |           | zero      |            | other    |          | <b>Total</b> |            |
|-------------------|-------------|----------|--------------|-----------|-------------|-----------|-----------|------------|----------|----------|--------------|------------|
|                   | %           | N        | %            | N         | %           | N         | %         | N          | %        | N        | %            | N          |
| <i>Definite</i>   | 1           | 1        | 27           | 49        | 27          | 49        | 45        | 82         | 1        | 2        | <b>62</b>    | <b>183</b> |
| <i>Indefinite</i> | 3           | 3        | 27           | 30        | 34          | 38        | 35        | 39         | 2        | 2        | <b>38</b>    | <b>112</b> |
| <i>Ambiguous</i>  | 0           | 0        | 0            | 0         | 50          | 1         | 50        | 1          | 0        | 0        | <b>1</b>     | <b>2</b>   |
| <b>Total</b>      | <b>1</b>    | <b>4</b> | <b>27</b>    | <b>79</b> | <b>30</b>   | <b>88</b> | <b>41</b> | <b>122</b> | <b>1</b> | <b>4</b> | <b>100</b>   | <b>297</b> |

Table 24: Contribution of definiteness of the antecedent to the selection of variants in non-subject relatives

#### 3.4.4.6 Sex

In non-subject relatives, the only notable difference between male and female speakers is in the lower rates of zero by female speakers (33%). *That* and *which* show only minor variation across male and female speakers.

| Factors      | <i>whom</i> |          | <i>which</i> |           | <i>that</i> |           | zero      |            | other    |          | <b>Total</b> |            |
|--------------|-------------|----------|--------------|-----------|-------------|-----------|-----------|------------|----------|----------|--------------|------------|
|              | %           | N        | %            | N         | %           | N         | %         | N          | %        | N        | %            | N          |
| male         | 1           | 2        | 25           | 40        | 27          | 44        | 46        | 74         | 1        | 1        | <b>54</b>    | <b>161</b> |
| female       | 0           | 0        | 32           | 32        | 32          | 32        | 33        | 33         | 3        | 3        | <b>34</b>    | <b>100</b> |
| unknown      | 6           | 2        | 19           | 7         | 33          | 12        | 42        | 15         | 0        | 0        | <b>12</b>    | <b>36</b>  |
| <b>Total</b> | <b>1</b>    | <b>4</b> | <b>27</b>    | <b>79</b> | <b>30</b>   | <b>88</b> | <b>41</b> | <b>122</b> | <b>1</b> | <b>4</b> | <b>100</b>   | <b>297</b> |

Table 25: Contribution of speaker sex to the selection of variants in non-subject relatives

The data from the *Salem Witchcraft Papers* suggests that there is a (relatively weak) sex effect for non-subject relatives – however, it is in the opposite direction to that reported by Huber (2017), who finds that female speakers use the zero relative more frequently (29% in the first period) than male speakers (19% in the first period), who predominate in the use of *wh*-relative markers. In the data reported in Table 25, female speakers lead in the use of *which* (32% compared to 25% for male speakers), while male speakers use the zero relative more frequently (46% compared to 33%). However, it is difficult to make a direct comparison as Huber (2017) does not separate subject and non-subject relatives when discussing sex. As a social variable, the significance and influence of gender roles is related to the specific society and context under investigation (Romaine, 2003: 109). There is very limited information available about the role of

women in 17<sup>th</sup> century colonial American society, and so any claims that might be made must perforce be cautious and tentative. Given that the variation is only really seen with zero non-subject relatives, and that it is only a minor difference, it would be perhaps unwise to claim that the relative system in 17<sup>th</sup> century Massachusetts is subject to sex effects. Such a simplistic explanation almost certainly does not account for the complex web of social relationships and interactions that are extant in any given community (D’Arcy & Tagliamonte, 2010), and as the differences between male and female speakers are still relatively slight, sex is likely not a particularly strong determinant of relative marker selection.

#### 3.4.4.7 Birthplace

Speaker birthplace does not appear to exert any influence over relativiser selection in non-subject relatives, again with the caveat that there are very low token counts for speakers born outside of New England. Speakers from New England and speakers from Britain show rates of *which* (30% and 29% respectively), *that* (29% each) and the zero relative (41% and 42% respectively) which are in line with the overall distribution of non-subject relatives.

| Factors             | <i>whom</i> |          | <i>which</i> |           | <i>that</i> |           | zero      |            | other    |          | <b>Total</b> |            |
|---------------------|-------------|----------|--------------|-----------|-------------|-----------|-----------|------------|----------|----------|--------------|------------|
|                     | %           | N        | %            | N         | %           | N         | %         | N          | %        | N        | %            | N          |
| New England         | 0           | 0        | 30           | 47        | 29          | 46        | 41        | 65         | 1        | 1        | <b>54</b>    | <b>159</b> |
| Britain             | 0           | 0        | 29           | 7         | 29          | 7         | 42        | 10         | 0        | 0        | <b>8</b>     | <b>24</b>  |
| Other <sup>12</sup> | 0           | 0        | 0            | 0         | 40          | 2         | 20        | 1          | 40       | 2        | <b>2</b>     | <b>5</b>   |
| Unknown             | 4           | 4        | 23           | 25        | 30          | 33        | 42        | 46         | 1        | 1        | <b>37</b>    | <b>109</b> |
| <b>Total</b>        | <b>1</b>    | <b>4</b> | <b>27</b>    | <b>79</b> | <b>30</b>   | <b>88</b> | <b>41</b> | <b>122</b> | <b>1</b> | <b>4</b> | <b>100</b>   | <b>297</b> |

Table 26: Contribution of speaker birthplace to the selection of variants in non-subject relatives

#### 3.4.4.8 Age

The effect of speaker age on variant selection in non-subject relatives patterns very similarly to subject relatives. Again, *which* is favoured among the youngest group of speakers (43%), slightly favoured

<sup>12</sup> Includes one speaker from South America, and one from Barbados. As there are very limited tokens (and a single speaker from each area), these have been subsumed under the heading ‘Other’.

among the middle group (33%), and is disfavoured by the oldest group of speakers (18%). Contrastively, zero is favoured among the older speakers (46%), has rates in line with overall distribution among the middle-aged speakers (39%), and is disfavoured by younger speakers (33%). *That* is fairly consistent, but as with subject relatives, does seem to be favoured by older speaker groups. *Whom* is extremely rare (only 4 tokens), and is not used in any appreciable way by the speakers. The data here points to a rise in *which* at the expense of zero, and to a lesser extent *that*.

| Factors      | <i>whom</i> |          | <i>which</i> |           | <i>that</i> |           | zero      |            | other    |          | <b>Total</b> |            |
|--------------|-------------|----------|--------------|-----------|-------------|-----------|-----------|------------|----------|----------|--------------|------------|
|              | %           | N        | %            | N         | %           | N         | %         | N          | %        | N        | %            | N          |
| Younger      | 0           | 0        | 43           | 26        | 25          | 15        | 33        | 20         | 0        | 0        | 21           | 61         |
| Middle       | 0           | 0        | 33           | 25        | 28          | 21        | 39        | 29         | 0        | 0        | 25           | 75         |
| Older        | 1           | 1        | 18           | 16        | 33          | 30        | 46        | 42         | 2        | 2        | 31           | 91         |
| Unknown      | 3           | 3        | 17           | 12        | 31          | 22        | 44        | 31         | 3        | 2        | 24           | 70         |
| <b>Total</b> | <b>1</b>    | <b>4</b> | <b>27</b>    | <b>79</b> | <b>30</b>   | <b>88</b> | <b>41</b> | <b>122</b> | <b>1</b> | <b>4</b> | <b>100</b>   | <b>297</b> |

Table 27: Contribution of speaker age to the selection of variants in non-subject relatives

The results for age with both subject and non-subject relatives, seem to indicate that there may be change occurring in the relative system in 17<sup>th</sup> century Massachusetts, namely an increase in *which*, and a subsequent decrease in zero and *that*. *Who/whom* are so rare in the data that it is not possible to make any firm conclusions about changes in their use. Equivalent data from periods earlier and later than the *Salem Witch Trials* would be required to lend more credence to the notion that *which* is increasing at the expense of *that* and zero, but this remains an interesting avenue to pursue in future research.

#### 3.4.4.9 Social Status

The distribution of non-subject relatives by position in the social hierarchy is relatively similar to subject relatives. *That* is again fairly stable across groupings (around 30%). WH is favoured among speakers in Group C (49%), and disfavoured by Group A (14%). Contrastively, zero is favoured among Group A (58%), and disfavoured by Group C (20%). Group B shows less susceptibility, slightly favouring the use of WH (32%) and slightly disfavours the use of *that* (24%).

| Factors        | WH        |           | <i>that</i> |           | zero      |            | other    |          | Total      |            |
|----------------|-----------|-----------|-------------|-----------|-----------|------------|----------|----------|------------|------------|
|                | %         | N         | %           | N         | %         | N          | %        | N        | %          | N          |
| <i>Group A</i> | 14        | 8         | 26          | 15        | 58        | 33         | 2        | 1        | 19         | 57         |
| <i>Group B</i> | 32        | 23        | 24          | 17        | 42        | 30         | 1        | 1        | 24         | 71         |
| <i>Group C</i> | 49        | 20        | 32          | 13        | 20        | 8          | 0        | 0        | 14         | 41         |
| <i>Group D</i> | 0         | 0         | 33          | 2         | 33        | 2          | 33       | 2        | 2          | 6          |
| <i>Unknown</i> | 26        | 32        | 34          | 41        | 40        | 49         | 0        | 0        | 41         | 122        |
| <b>Total</b>   | <b>28</b> | <b>83</b> | <b>30</b>   | <b>88</b> | <b>41</b> | <b>122</b> | <b>1</b> | <b>4</b> | <b>100</b> | <b>297</b> |

Table 28: Contribution of speaker social status to the selection of variants in non-subject relatives

It was hypothesised that speakers in Group A, representing the upper echelon of society, would use the more prestigious WH forms most frequently, and the more colloquial zero relative least frequently. The data instead shows a fairly marked difference in the use of WH and zero by speakers in Groups A and C. It is not entirely clear why this should be the case, but token counts are relatively low for each grouping in the hierarchy.

Turning to the distribution of speaker groups by their role in the trial, Group B are most often the accused (49%), while Group A are usually officials (61%), and Group C are accusers (46%). This pattern therefore holds true across subject and non-subject relatives. Examining the distribution of variants by speaker role (Table 29), it can be seen that the effect of speaker role is stronger among non-subject relatives. Accusers strongly favour the use of WH (49%), slightly disfavours the use of *that* and zero (21% and 31% respectively), though again there are small token numbers for each speaker role. Officials disfavour the use of WH (17%), and favour the use of zero (49%). The accused and witnesses show rates roughly in line with overall distribution.

| Factors        | WH |    | <i>that</i> |    | zero |    | other |   | Total |    |
|----------------|----|----|-------------|----|------|----|-------|---|-------|----|
|                | %  | N  | %           | N  | %    | N  | %     | N | %     | N  |
| <i>Accused</i> | 20 | 15 | 36          | 27 | 41   | 31 | 4     | 3 | 26    | 76 |
| <i>Accuser</i> | 49 | 19 | 21          | 8  | 31   | 12 | 0     | 0 | 13    | 39 |

|                 |           |           |           |           |           |            |          |          |            |            |
|-----------------|-----------|-----------|-----------|-----------|-----------|------------|----------|----------|------------|------------|
| <i>Official</i> | 17        | 7         | 32        | 13        | 49        | 20         | 2        | 1        | 14         | 41         |
| <i>Witness</i>  | 33        | 30        | 27        | 25        | 40        | 36         | 0        | 0        | 31         | 91         |
| <i>Unknown</i>  | 24        | 12        | 30        | 15        | 46        | 23         | 0        | 0        | 17         | 50         |
| <b>Total</b>    | <b>28</b> | <b>83</b> | <b>30</b> | <b>88</b> | <b>41</b> | <b>122</b> | <b>1</b> | <b>4</b> | <b>100</b> | <b>297</b> |

Table 29: Contribution of speaker role to the selection of variants in non-subject relatives

#### 3.4.4.10 Document Type

In non-subject relative clauses, the effects of document type are less clear. In examinations, consistent with the rates in subject relatives, the *wh*-relativisers are disfavoured (11%). *That* (38%) and zero (49%) are favoured in this context. This is consistent with the association of the zero relative with the spoken language, and the *wh*-relatives with the written language, as examinations are the most speech-like of the surviving texts (Huber, 2017; Romaine, 1980: 234; Tagliamonte, 2013: 95). The distribution of variants in letters is consistent with the overall distribution in non-subject relatives. There are only 4 non-subject relatives in official documents, too few for any conclusive statements, but it is worth noting the absence of *that* and zero in this context.

It is in depositions that an unexpected pattern emerges – the *wh*-relatives (predominantly *which*, accounting for N=55/56 tokens) are the most frequent variant. *That* (23%) and zero (35%) are both disfavoured in non-subject relatives in this text type. What is observable is that non-subject relatives in depositions are dominated by inanimate and noun phrase antecedents, both contexts which favour the use of *which*. Of the 133 tokens, 117 (88%) have inanimate antecedents, and 119 (89%) have noun phrase antecedents. One possible explanation is that the predominance of these types of antecedent is obscuring the effects of text type.

| Factors             | WH |    | <i>that</i> |    | zero |    | other |   | Total |     |
|---------------------|----|----|-------------|----|------|----|-------|---|-------|-----|
|                     | %  | N  | %           | N  | %    | N  | %     | N | %     | N   |
| <i>Examinations</i> | 11 | 13 | 38          | 45 | 49   | 58 | 2     | 2 | 40    | 118 |
| <i>Depositions</i>  | 42 | 56 | 23          | 30 | 35   | 47 | 0     | 0 | 45    | 133 |
| <i>Letters</i>      | 26 | 11 | 31          | 13 | 40   | 17 | 2     | 1 | 14    | 42  |



|                           |           |           |           |           |           |            |          |          |            |            |
|---------------------------|-----------|-----------|-----------|-----------|-----------|------------|----------|----------|------------|------------|
| <i>Official Documents</i> | 75        | 3         | 0         | 0         | 0         | 0          | 25       | 1        | 1          | 4          |
| <b>Total</b>              | <b>28</b> | <b>83</b> | <b>30</b> | <b>88</b> | <b>41</b> | <b>122</b> | <b>1</b> | <b>4</b> | <b>100</b> | <b>297</b> |

Table 30: Contribution of document type to the selection of variants in non-subject relatives

Discussion will now turn to the statistical significance of the factor groups tested.

### 3.4.5 Statistical Analysis

In order to test the statistical significance of each of the internal and external factor groups investigated here, a mixed-effects regression analysis was conducted using Rbrul (Johnson, 2009). Individual (i.e. the people to whom each token is attributed) was run as a random effect to account for the potential effects of inter-speaker variation. Subject and non-subject relatives were analysed separately. Factor groups that are not statistically significant are indicated with **(n.s.)**. In certain cases, low token counts necessitated the consolidation of certain factors. Thus, matrix clause construction type, *it*-clefts, existentials, and stative possessives are grouped together as propositionally light main verbs. Similarly, in non-subject relatives, clause length is operationalised as a binary distinction between short relative clauses, and longer relative clauses (combining the medium-length and long clauses). The effects of preposition stranding are not subjected to statistical analysis due to the almost categorical absence of *wh*-relatives in preposition stranding contexts.

First, some explanation of the contents of the tables is required. The overall probability that the target variant will occur in the data is indicated by the input probability. The log likelihood represents how well the regression model fits the dataset. The proportion of variation explained by the model is shown by the R2 total. Interpretation of the table relies on two key lines of evidence: statistical significance of the influence of factor groups on the choice of variants, and the constraint hierarchy. The constraint hierarchy is determined by the ordering of factors within each factor group, based on their relative impact on the choice of variants. This is represented by the log odds and centred factor weights (FW). A log odds value with a positive value is indicative of a favouring effect on variant choice, while a negative value is disfavouring. Similarly, a factor weight over 0.5 is evidence of a factor that favours the variant under investigation, and a factor weight under 0.5 disfavors. Considered together, statistical significance and the direction of effects (i.e. the ordering of probability co-efficients from largest to smallest in each

predictor) give a detailed insight into the linguistic and social conditioning of the variable in the *Salem Witchcraft Papers*.

#### 3.4.5.1 Subject Relatives

In subject relatives, definiteness and the length of the relative clause are statistically significant predictors of the use of *that*. In particular, *that* is favoured in shorter and medium-length relative clauses, and with definite antecedents. Longer relative clauses and indefinite antecedents disfavour the use of *that*. All remaining factor groups are non-significant.

|                                                            |          |            |    |     |
|------------------------------------------------------------|----------|------------|----|-----|
| Input probability                                          | 0.676    |            |    |     |
| Log likelihood                                             | -238.356 |            |    |     |
| R2 total                                                   | 0.211    |            |    |     |
| N                                                          | 441      |            |    |     |
|                                                            |          |            |    |     |
| <b>Length of<br/>Relative Clause<br/>(p &lt; 0.000329)</b> | Log Odds | Centred FW | %  | N   |
| Short                                                      | 0.509    | 0.625      | 77 | 282 |
| Medium                                                     | 0.465    | 0.614      | 76 | 116 |
| Long                                                       | -0.974   | 0.274      | 42 | 43  |
| <b>Definiteness (p &lt;<br/>0.000315)</b>                  |          |            |    |     |
| Definite                                                   | 0.438    | 0.608      | 81 | 252 |
| Indefinite                                                 | -0.438   | 0.392      | 62 | 189 |
| <b>Adjacency (n.s.)</b>                                    |          |            |    |     |
| Adjacent                                                   | 0.22     | 0.555      | 74 | 390 |
| Non-Adjacent                                               | -0.22    | 0.445      | 65 | 51  |
| <b>Animacy (n.s.)</b>                                      |          |            |    |     |

|                                               |        |       |    |     |
|-----------------------------------------------|--------|-------|----|-----|
| Human                                         | 0.054  | 0.514 | 75 | 294 |
| Inanimate                                     | -0.054 | 0.486 | 70 | 147 |
| <b>Matrix Clause Construction Type (n.s.)</b> |        |       |    |     |
| Propositionally Light Main Clauses            | 0.038  | 0.51  | 76 | 101 |
| Other                                         | -0.038 | 0.49  | 72 | 340 |
| <b>Antecedent Type (n.s.)</b>                 |        |       |    |     |
| Pronoun                                       | 0.052  | 0.513 | 78 | 111 |
| Noun Phrase                                   | -0.052 | 0.487 | 72 | 330 |

Table 31: Rbrul mixed-effects logistic regression analysis of the contribution of linguistic factors to the selection of relativiser *that* in subject relatives

Of the social factor groups tested, document type alone exerted a significant effect on the use of *that*, which is favoured in examinations and depositions, and disfavoured in letters. This is suggestive of an association of *that* with less formal, more speech-like contexts. Official documents were too infrequent for inclusion in the regression analysis. While not statistically significant, it is perhaps worth noting that *that* is most frequent among older speakers, as would be expected if the *wh*-relatives are gaining ground at the expense of *that*.

|                   |          |  |  |  |
|-------------------|----------|--|--|--|
| Input probability | 0.715    |  |  |  |
| Log likelihood    | -240.714 |  |  |  |
| R2 total          | 0.097    |  |  |  |
| N                 | 424      |  |  |  |
|                   |          |  |  |  |

| <b>Document Type<br/>(p &lt; 0.0484)</b> | Log Odds | Centred FW | %  | N   |
|------------------------------------------|----------|------------|----|-----|
| Examinations                             | 0.414    | 0.602      | 79 | 176 |
| Depositions                              | 0.102    | 0.525      | 72 | 191 |
| Letters                                  | -0.516   | 0.374      | 60 | 57  |
| <b>Sex (n.s.)</b>                        |          |            |    |     |
| Female                                   | 0.142    | 0.535      | 75 | 178 |
| Male                                     | -0.142   | 0.465      | 70 | 208 |
| <b>Age (n.s.)</b>                        |          |            |    |     |
| Older                                    | 0.312    | 0.577      | 79 | 104 |
| Middle                                   | -0.019   | 0.495      | 69 | 134 |
| Younger                                  | -0.293   | 0.427      | 68 | 104 |
| <b>Birthplace (n.s.)</b>                 |          |            |    |     |
| Britain                                  | 0.352    | 0.587      | 81 | 43  |
| America                                  | -0.352   | 0.413      | 71 | 237 |

Table 32: Rbrul mixed-effects logistic regression analysis of the contribution of social factors to the selection of relativiser that in subject relatives

Turning now to the statistical contribution of factor groups to the use of *which* in subject relatives, it can be seen in Table 33 that animacy, the length of the relative clause, and the matrix clause construction type are significant. The favouring of *which* with inanimate antecedents lends credence to Ball's (1996: 252) claim that the 17<sup>th</sup> century saw the emergence of an animacy parameter. *Which* is also favoured in longer clauses, consistent with the association of overt relative markers (e.g. the *wh*-relatives) with more cognitively complex constructions (Rohdenburg, 1996). This is additionally supported by the (non-significant) preference for *which* in non-adjacent contexts.

|                   |          |  |  |  |
|-------------------|----------|--|--|--|
| Input probability | 0.083    |  |  |  |
| Log likelihood    | -143.595 |  |  |  |
| R2 total          | 0.577    |  |  |  |
| N                 | 441      |  |  |  |

| <b>Animacy (p &lt; 5.89e-06)</b>                       | Log Odds | Centred FW | %  | N   |
|--------------------------------------------------------|----------|------------|----|-----|
| Inanimate                                              | 0.912    | 0.713      | 22 | 147 |
| Human                                                  | -0.912   | 0.287      | 9  | 294 |
| <b>Length of Relative Clause (p &lt; 6.82e-07)</b>     |          |            |    |     |
| Long                                                   | 1.587    | 0.83       | 44 | 43  |
| Medium                                                 | -0.434   | 0.393      | 14 | 116 |
| Short                                                  | -1.152   | 0.24       | 9  | 282 |
| <b>Matrix Clause Construction Type (p &lt; 0.0384)</b> |          |            |    |     |
| Other                                                  | 0.536    | 0.631      | 15 | 340 |
| Propositionally Light Main Clauses                     | -0.536   | 0.369      | 7  | 101 |
| <b>Definiteness (n.s.)</b>                             |          |            |    |     |
| Indefinite                                             | 0.392    | 0.597      | 19 | 189 |
| Definite                                               | -0.392   | 0.403      | 10 | 252 |
| <b>Adjacency (n.s.)</b>                                |          |            |    |     |
| Non-Adjacent                                           | 0.281    | 0.57       | 20 | 51  |
| Adjacent                                               | -0.281   | 0.43       | 13 | 390 |
| <b>Grammatical Category of Antecedent (n.s.)</b>       |          |            |    |     |
| Noun Phrase                                            | 0.202    | 0.55       | 16 | 330 |
| Pronoun                                                | -0.202   | 0.45       | 7  | 111 |

Table 33: Rbrul mixed-effects logistic regression analysis of the contribution of linguistic factors to the selection of relativiser which in subject relatives

As Table 34 shows, the data from the *Salem Witchcraft Papers* supports the idea that the *wh*-relatives (in this case *which*) are increasing in frequency at the end of the 17<sup>th</sup> century (Fowler & Fowler, 1931: 80). *Which* is favoured among the youngest speaker group, and disfavoured among older speakers, as would be expected of a change in progress (supported by the non-significant favouring of *which* with female speakers). However, *which* does not yet appear to have infiltrated speech-related genres, as it is

disfavoured in examinations. This supports Romaine's (1982: 212) claim that the *wh*-relatives in Early Modern English have only established a secure foothold in canonical written genres.

|                                            |          |            |    |     |
|--------------------------------------------|----------|------------|----|-----|
| Input probability                          | 0.117    |            |    |     |
| Log likelihood                             | -158.277 |            |    |     |
| R2 total                                   | 0.258    |            |    |     |
| N                                          | 424      |            |    |     |
|                                            |          |            |    |     |
| <b>Document Type<br/>(p &lt; 0.000429)</b> | Log Odds | Centred FW | %  | N   |
| Depositions                                | 0.583    | 0.642      | 20 | 191 |
| Letters                                    | 0.419    | 0.603      | 16 | 57  |
| Examinations                               | -1.001   | 0.269      | 5  | 176 |
| <b>Age (p &lt; 0.0121)</b>                 |          |            |    |     |
| Younger                                    | 0.536    | 0.631      | 22 | 104 |
| Middle                                     | -0.027   | 0.493      | 14 | 134 |
| Older                                      | -0.510   | 0.375      | 9  | 104 |
| <b>Birthplace (n.s.)</b>                   |          |            |    |     |
| America                                    | 0.695    | 0.667      | 15 | 237 |
| Britain                                    | -0.695   | 0.333      | 5  | 43  |
| <b>Sex (n.s.)</b>                          |          |            |    |     |
| Female                                     | 0.121    | 0.53       | 16 | 178 |
| Male                                       | -0.121   | 0.47       | 12 | 208 |

Table 34: Rbrul mixed-effects logistic regression analysis of the contribution of social factors to the selection of relativiser which in subject relatives

### 3.4.5.2 Non-Subject Relatives

In non-subject relatives, matrix clause construction type is the only linguistic factor group to significantly contribute to the use of *that*, which is favoured with propositionally light main verbs (*it*-clefts, existentials, stative possessives).

|                                   |          |            |   |   |
|-----------------------------------|----------|------------|---|---|
| Input probability                 | 0.353    |            |   |   |
| Log likelihood                    | -179.83  |            |   |   |
| R2 total                          | 0.02     |            |   |   |
| N                                 | 302      |            |   |   |
|                                   |          |            |   |   |
| <b>Matrix Clause Construction</b> | Log Odds | Centred FW | % | N |

|                                         |        |       |    |     |
|-----------------------------------------|--------|-------|----|-----|
| <b>Type (p &lt; 0.0286)</b>             |        |       |    |     |
| Propositionally Light Main Verb         | 0.394  | 0.597 | 45 | 38  |
| Other                                   | -0.394 | 0.403 | 27 | 264 |
| <b>Definiteness (n.s.)</b>              |        |       |    |     |
| Indefinite                              | 0.178  | 0.544 | 34 | 113 |
| Definite                                | -0.178 | 0.456 | 26 | 187 |
| <b>Antecedent Type (n.s.)</b>           |        |       |    |     |
| Pronoun                                 | 0.175  | 0.544 | 36 | 53  |
| Noun Phrase                             | -0.175 | 0.456 | 28 | 249 |
| <b>Adjacency (n.s.)</b>                 |        |       |    |     |
| Adjacent                                | 0.209  | 0.552 | 30 | 267 |
| Non-Adjacent                            | -0.209 | 0.448 | 23 | 35  |
| <b>Animacy (n.s.)</b>                   |        |       |    |     |
| Human                                   | 0.048  | 0.512 | 32 | 56  |
| Inanimate                               | -0.048 | 0.488 | 29 | 246 |
| <b>Length of relative clause (n.s.)</b> |        |       |    |     |
| Longer                                  | 0.006  | 0.502 | 30 | 81  |
| Shorter                                 | -0.006 | 0.498 | 29 | 221 |

Table 35: Rbrul mixed-effects logistic regression analysis of the contribution of linguistic factors to the selection of relativiser that in non-subject relatives

Of the social factor groups tested, document type was statistically significant, with *that* being favoured in examinations, and disfavoured in depositions. Though not significant, it is of note that *that* is favoured among older and male speakers, consistent with it being an older and more conservative relative marker.

|                                     |          |            |    |     |
|-------------------------------------|----------|------------|----|-----|
| Input probability                   | 0.296    |            |    |     |
| Log likelihood                      | -173.657 |            |    |     |
| R2 total                            | 0.036    |            |    |     |
| N                                   | 293      |            |    |     |
|                                     |          |            |    |     |
| <b>Document Type (p &lt; 0.025)</b> | Log Odds | Centred FW | %  | N   |
| Examinations                        | 0.365    | 0.59       | 38 | 114 |

|                          |        |       |    |     |
|--------------------------|--------|-------|----|-----|
| Letters                  | 0.030  | 0.508 | 30 | 43  |
| Depositions              | -0.396 | 0.402 | 22 | 136 |
| <b>Age (n.s.)</b>        |        |       |    |     |
| Older                    | 0.259  | 0.564 | 34 | 89  |
| Middle                   | -0.010 | 0.498 | 28 | 75  |
| Younger                  | -0.249 | 0.438 | 23 | 64  |
| <b>Sex (n.s.)</b>        |        |       |    |     |
| Male                     | 0.055  | 0.514 | 30 | 99  |
| Female                   | -0.055 | 0.486 | 28 | 157 |
| <b>Birthplace (n.s.)</b> |        |       |    |     |
| Britain                  | 0.045  | 0.511 | 29 | 24  |
| America                  | -0.045 | 0.489 | 28 | 162 |

Table 36: Rbrul mixed-effects logistic regression analysis of the contribution of social factors to the selection of relativiser that in non-subject relatives

Comparing Table 37 below with Table 33 above, it can be observed that the linguistic conditioning of relative marker *which* is consistent across subject and non-subject relatives. In both cases, animacy, clause length, and matrix clause construction type are statistically significant factor groups, favouring the use of *which* with inanimate antecedents, longer clauses, and non-propositionally light main verbs respectively. The key difference is the statistical significance of adjacency in non-subject relative clauses, though the ranking of constraints is the same, favouring *which* in non-adjacent contexts. This finding is consistent with the association of *which* with more cognitively complex environments, and perhaps speaks to the greater influence of processing constraints in non-subject relatives. It is worth highlighting here the statistical significance of animacy as a factor group conditioning the selection of *which* in both subject (Table 33) and non-subject (Table 37) relatives. This is robust evidence of the animacy constraint proposed by Ball (1996), restricting *which* to inanimate antecedents.

|                                 |          |            |    |     |
|---------------------------------|----------|------------|----|-----|
| Input probability               | 0.205    |            |    |     |
| Log likelihood                  | -143.035 |            |    |     |
| R2 total                        | 0.448    |            |    |     |
| N                               | 300      |            |    |     |
|                                 |          |            |    |     |
| <b>Animacy (p &lt; 0.00365)</b> | Log Odds | Centred FW | %  | N   |
| Inanimate                       | 0.72     | 0.673      | 30 | 245 |
| Human                           | -0.72    | 0.327      | 11 | 55  |



|                                                         |        |       |    |     |
|---------------------------------------------------------|--------|-------|----|-----|
| <b>Length of Relative Clause (p &lt; 0.00942)</b>       |        |       |    |     |
| Longer                                                  | 0.467  | 0.615 | 37 | 81  |
| Shorter                                                 | -0.467 | 0.385 | 22 | 219 |
| <b>Adjacency (p &lt; 1.11e-07)</b>                      |        |       |    |     |
| Non-Adjacent                                            | 1.247  | 0.776 | 66 | 35  |
| Adjacent                                                | -1.247 | 0.223 | 21 | 265 |
| <b>Matrix Clause Construction Type (p &lt; 0.00117)</b> |        |       |    |     |
| Other                                                   | 1.023  | 0.735 | 29 | 262 |
| Propositionally Light Main Verb                         | -1.023 | 0.265 | 8  | 38  |
| <b>Antecedent Type (n.s.)</b>                           |        |       |    |     |
| Noun Phrase                                             | 0.211  | 0.553 | 28 | 247 |
| Pronoun                                                 | -0.211 | 0.447 | 21 | 53  |
| <b>Definiteness (n.s.)</b>                              |        |       |    |     |
| Indefinite                                              | 0.041  | 0.51  | 27 | 113 |
| Definite                                                | -0.041 | 0.49  | 26 | 187 |

Table 37: Rbrul mixed-effects logistic regression analysis of the contribution of linguistic factors to the selection of relativiser which in non-subject relatives

Turning to social factor groups, it can be observed that again, document type is a statistically significant predictor of the use of *which*, which is favoured in depositions and letters, and disfavoured in examinations. *Which* is also favoured among speakers who were born in America. Consistent with a change in progress (though not statistically significant), *which* is favoured with younger and female speakers (Labov, 1990).

|                   |          |  |  |  |
|-------------------|----------|--|--|--|
| Input probability | 0.226    |  |  |  |
| Log likelihood    | -146.705 |  |  |  |
| R2 total          | 0.468    |  |  |  |
| N                 | 291      |  |  |  |
|                   |          |  |  |  |

| <b>Document Type<br/>(p &lt; 6.85e-08)</b> | Log Odds | Centred FW | %  | N   |
|--------------------------------------------|----------|------------|----|-----|
| Depositions                                | 1.119    | 0.754      | 40 | 136 |
| Letters                                    | 0.226    | 0.556      | 23 | 43  |
| Examinations                               | -1.345   | 0.207      | 11 | 112 |
| <b>Birthplace (p &lt; 0.0127)</b>          |          |            |    |     |
| America                                    | 0.242    | 0.56       | 29 | 160 |
| Britain                                    | -0.242   | 0.44       | 29 | 24  |
| <b>Age (n.s.)</b>                          |          |            |    |     |
| Younger                                    | 0.602    | 0.646      | 41 | 63  |
| Middle                                     | -0.026   | 0.494      | 32 | 75  |
| Older                                      | -0.576   | 0.36       | 18 | 88  |
| <b>Sex (n.s.)</b>                          |          |            |    |     |
| Female                                     | 0.339    | 0.584      | 33 | 98  |
| Male                                       | -0.339   | 0.416      | 24 | 156 |

Table 38: Rbrul mixed-effects logistic regression analysis of the contribution of social factors to the selection of relativiser *which* in non-subject relatives

In non-subject relatives, processing constraints make a significant contribution to the selection of the zero relative (Quirk, 1957; Rohdenburg, 1996; Tottie & Johansson, 2015). In particular, less cognitively complex environments like shorter relative clauses and adjacency between the relative marker and the antecedent, favour the use of zero. In each factor group, the direction of constraints is the reverse of the situation for *which* (Table 37 above). This suggests that in non-subject relatives, the zero relative operates in direct contrast to relative marker *which*. *That* appears to occupy middle ground: in some factor groups (e.g. definiteness, relative clause length) patterning like *which*, and in others patterning like the zero relative (e.g. matrix clause construction type, adjacency, animacy, antecedent type).

|                                                           |          |            |    |     |
|-----------------------------------------------------------|----------|------------|----|-----|
| Input probability                                         | 0.104    |            |    |     |
| Log likelihood                                            | -183.012 |            |    |     |
| R2 total                                                  | 0.313    |            |    |     |
| N                                                         | 300      |            |    |     |
|                                                           |          |            |    |     |
| <b>Length of<br/>Relative Clause<br/>(p &lt; 0.00566)</b> | Log Odds | Centred FW | %  | N   |
| Shorter                                                   | 0.406    | 0.6        | 45 | 219 |

|                                               |        |       |    |     |
|-----------------------------------------------|--------|-------|----|-----|
| Longer                                        | -0.406 | 0.4   | 28 | 81  |
| <b>Adjacency (p &lt; 3.04e-08)</b>            |        |       |    |     |
| Adjacent                                      | 1.716  | 0.848 | 45 | 265 |
| Non-Adjacent                                  | -1.716 | 0.152 | 3  | 35  |
| <b>Definiteness (n.s.)</b>                    |        |       |    |     |
| Definite                                      | 0.163  | 0.541 | 44 | 187 |
| Indefinite                                    | -0.163 | 0.459 | 35 | 113 |
| <b>Matrix Clause Construction Type (n.s.)</b> |        |       |    |     |
| Propositionally Light Main Clause             | 0.14   | 0.535 | 42 | 38  |
| Other                                         | -0.14  | 0.465 | 40 | 262 |
| <b>Antecedent Type (n.s.)</b>                 |        |       |    |     |
| Pronoun                                       | 0.031  | 0.508 | 42 | 53  |
| Noun Phrase                                   | -0.031 | 0.492 | 40 | 247 |
| <b>Animacy (n.s.)</b>                         |        |       |    |     |
| Human                                         | 0.01   | 0.502 | 42 | 55  |
| Inanimate                                     | -0.01  | 0.498 | 40 | 245 |

Table 39: Rbrul mixed-effects logistic regression analysis of the contribution of linguistic factors to the selection of the zero relative in non-subject relatives

As Table 40 demonstrates, the non-subject zero relative is favoured in more speech-like text types (examinations), and among male speakers. While not statistically significant, it is perhaps worth noting the preference for the zero relative among older and middle-aged speakers.

|                                      |          |            |    |     |
|--------------------------------------|----------|------------|----|-----|
| Input probability                    | 0.395    |            |    |     |
| Log likelihood                       | -189.997 |            |    |     |
| R2 total                             | 0.063    |            |    |     |
| N                                    | 291      |            |    |     |
|                                      |          |            |    |     |
| <b>Document Type (p &lt; 0.0091)</b> | Log Odds | Centred FW | %  | N   |
| Examinations                         | 0.462    | 0.613      | 50 | 112 |
| Letters                              | -0.085   | 0.479      | 40 | 43  |
| Depositions                          | -0.377   | 0.407      | 35 | 136 |
| <b>Sex (p &lt; 0.0063)</b>           |          |            |    |     |

|                          |        |       |    |     |
|--------------------------|--------|-------|----|-----|
| Male                     | 0.394  | 0.597 | 47 | 156 |
| Female                   | -0.394 | 0.403 | 33 | 98  |
| <b>Age (n.s.)</b>        |        |       |    |     |
| Older                    | 0.311  | 0.577 | 48 | 88  |
| Middle                   | 0.059  | 0.515 | 39 | 75  |
| Younger                  | -0.370 | 0.409 | 30 | 63  |
| <b>Birthplace (n.s.)</b> |        |       |    |     |
| Britain                  | 0.096  | 0.524 | 42 | 24  |
| America                  | -0.096 | 0.476 | 40 | 160 |

Table 40: Rbrul mixed-effects logistic regression analysis of the contribution of social factors to the selection of the zero relative in non-subject relatives

### 3.5 Discussion

3.5.1 Objective (i): to reconstruct the variable system of the target linguistic phenomenon as it exists in contemporaneous speech-related data

Romaine (1982) claims that the inventory of relative markers has remained essentially unchanged since the 17<sup>th</sup> century. While it is certainly the case that the same inventory of relative markers found in the 17<sup>th</sup> century is ostensibly available to present-day English speakers, usage data such as that reported in D'Arcy and Tagliamonte (2010) and Levey and Hill (2013) strongly suggests that there *have* been changes in the distribution and conditioning of variants. In modern spoken North American English varieties, the primary alternation is between *that* and *who* for subject relatives, and between *that* and the zero relative in non-subject relative clauses. This distribution of variants is in contrast to my results for 17<sup>th</sup> century Massachusetts, where *who* is marginal even in subject relatives with human antecedents, and *which* is the most frequently used *wh*-relative marker.

The distribution of variants across age groups is suggestive of a trend, in both subject and non-subject relatives, where *which* is increasing at the expense of *that* and the zero relative marker. Younger speakers consistently show higher rates of *which* than older speakers, who show higher rates of *that* and zero. It might then be the case that *which* is in the process of infiltrating the spoken language (*who* and *whom* are too rare to arrive at robust conclusions about the behaviour of either). If this were the case, then American English data from later periods may show higher rates of *which* (and consequently lower rates of *that* and zero). This does not, however, appear to be the case.

Levey and DeRooy (2021) report on the use of restrictive relatives in 19<sup>th</sup> century American English in the Corpus of American Civil War Letters. Most crucially, they report on relative usage by speakers from Massachusetts, allowing some degree of comparison with the data from the *Salem Witchcraft Papers*. There are some key differences and similarities in the data. For example, Levey and DeRooy (2021: 18) report higher rates of zero (17%), and consequently lower rates of *which* (9%) in subject position than is the case in the *Salem Witchcraft Papers*. *That* is slightly less frequent in subject position (66%) than in my data. In non-subject position, they report notably lower rates of *which* (8%), and higher rates of *that* (55%), while zero remains almost unchanged (36%). The results in Levey and DeRooy (2021) provisionally appear to suggest that by the 19<sup>th</sup> century, the use of *which* in American English had declined in favour of *that*.

The hypothesis that *which* experienced decline after the end of the 17<sup>th</sup> century is supported by present-day American English data reported by Hinrichs et al. (2015), who examined relativisation in modern written American English, comparing two corpora – the Brown Corpus of texts from 1961, and the Frown corpus of texts from 1992. The data from these corpora suggests a decline in the use of *which* in written American data – in Brown, *which* accounts for 46% of subject relatives and 18% of non-subject relatives, while in Frown *which* is selected 12% of the time for subject relatives and in only 4% of cases with non-subject relatives. The findings in Hinrichs et al. (2015) are congruent with those in modern North American *spoken* varieties, as detailed by D’Arcy and Tagliamonte (2010) and Levey and Hill (2013). D’Arcy and Tagliamonte (2010: 389), drawing on data from the Toronto English Archive, find that the *wh*-relatives are overwhelmingly (almost categorically) represented by *who*, accounting for 17.2% of variants. *Whose*, *which*, and *whom* are correspondingly infrequent, with their combined rates accounting for 0.5% of relatives in the data. Levey and Hill (2013: 47) reach similar conclusions about the status of the *wh*-relatives in Oshawa-Whitby (Poplack et al., 2006), where *who* is the only *wh*-relative to be used to any great extent – representing 13.1% of relative markers in the data. Again, *which* and *whose* are exceedingly rare, accounting for 0.5% and 0.4% of relative markers respectively. In both studies, *who* is found to be primarily restricted to subject relatives; and contrastively, the zero relative is almost exclusively restricted to non-subject contexts.

One of the major findings of this thesis concerns the statistically significant influence of internal processing constraints on the selection of relative markers. This is in effect in both subject and non-subject

relatives, where there is a marked preference for *wh*-relatives in more cognitively complex structures, and a preference for the zero relative (and to a lesser extent, *that*) in less cognitively complex structures. The length of the relative clause and adjacency between relative marker and antecedent were operationalised as indicators of complexity, working under the hypothesis that longer clauses and non-adjacency are more complex than shorter clauses and adjacency (and would therefore favour *wh*-relatives). The effects of processing constraints on relativiser choice have seen only limited attention in the literature (e.g. Quirk, 1957; Rohdenburg, 1996; Tottie & Johansson, 2015), so this result represents a notable finding.

Summarizing, the system of restrictive relativisation in 17<sup>th</sup> century Massachusetts English is best defined by three qualities: the dominance of *that* and the zero relative in more speech-like texts; indications of the animacy parameter constraining *who(m/se)* to human antecedents and *which* to non-human antecedents; and the significant contribution of processing constraints to the choice between *which* and the zero relative. Of the *wh*-relatives, only *which* is used to any great extent, and is predominantly restricted to contexts where *that* or the zero relative are disfavoured or constrained in some way.

3.5.2 Objective (ii): to refute, confirm, or refine existing claims in the literature regarding the target phenomenon

#### 3.5.2.1 Confirmation of previous claims

A major claim in the literature is Ball's (1996) claim that the 17<sup>th</sup> century witnesses the emergence of an animacy parameter, whereby *which* became confined to non-human antecedents. *That* is also claimed to gain increasing association with non-human antecedents. Statistical analysis (Section 3.4.5) provides robust evidence of the animacy constraint at the end of the 17<sup>th</sup> century, though the *wh*-relatives are only really operative in less speech-like text types. This is consistent with Ball (1996), who reports that by the end of the 17<sup>th</sup> century, the animacy constraint has only affected conventional written genres. In particular, *which* is statistically favoured with non-human antecedents. *Who* is exclusively found with human antecedents, and *whom* with human antecedents in all but one instance, where the antecedent is an animal. Relative marker *that* does not yet show the association with non-human antecedents that has been claimed for later periods (Ball, 1996), and is in fact marginally preferred with human antecedents. The results reported in this dissertation tentatively support the notion that the animacy constraint is not yet fully operative in speech-like varieties, and had not advanced far beyond conventional written styles.

### 3.5.2.2 Refinement of previous claims

One of the defining claims in the literature on relativisation in Early Modern English is Romaine's (1982: 212) assertion that while the *wh*-relatives have infiltrated the written language, the spoken language has hardly been affected. It is fairly clear that at the end of the 17<sup>th</sup> century, the *wh*-relatives have not infiltrated the spoken language (or at least speech-like texts) wholesale. *Who* and *whom* are extremely rare in the data, and are marginal even in their preferred niches (i.e. with human antecedents in subject and non-subject relatives respectively). *Which* is the only *wh*-relative to be used to any appreciable extent, accounting for 19% of total restrictive relative clauses in the data. Even so, statistical analysis demonstrates that *which* is disfavoured in examinations, the most speech-like text type in the dataset.

*Which* is the relative marker of choice in select contexts – in long clauses (10+ words) for both subject and non-subject relatives, in non-adjacent non-subject relatives, and in pied-piping structures. In each instance, the context that favours the selection of *which* is a minority. For example, pied-piping constructions appear 22% of the time in total compared to 78% stranding. The case is much the same with long clauses and non-adjacent contexts. There are only 35 non-adjacent non-subject relatives (where *which* accounts for 66% of tokens) compared to 262 adjacent non-subject relatives. There are 42 instances of long subject relative clauses compared to 282 short, and 18 counts of long non-subject relative clauses compared to 216 short. This pattern is strikingly consistent.

The contexts favouring the selection of *which* are confined to those where zero and *that* are constrained for one reason or another. Continuing with the example of preposition stranding, zero and *that* are not available in pied-piping contexts, and so it is perhaps unsurprising that *which* (the most frequent of the *wh*-relatives) dominates this context. Similarly, complexity-based processing constraints disfavour the selection of the zero relative (and to a lesser extent, *that*) in longer clauses and in non-adjacent contexts, which are exactly the contexts where *which* is favoured.

It is also possible to contribute to the literature regarding the effects of antecedent type and definiteness. As discussed previously (e.g. in Section 3.4.3.4), the literature is inconsistent on the role played by antecedent type and definiteness, with the zero relative being associated variously with pronominal antecedents (Huddleston, 1971; Piercey et al., 2011; Tottie, 1995; Tottie & Rey, 1997); indefinite noun phrase antecedents (Tagliamonte et al., 2005; Tottie & Harvie, 2000); and definite antecedents (Wasow

et al., 2011). My results suggest that in both subject and non-subject relatives, the zero relative is favoured with pronominal antecedents, while *that* is favoured with definite and pronominal antecedents. The *wh*-relatives, in contrast, are favoured with indefinite and nominal antecedents.

### 3.5.2.3 Refutation of previous claims

Huber (2017), looking at data from the *Old Bailey Corpus*, notes that female speakers lead in the use of the zero relative, while male speakers lead in the use of *wh*-relatives. This claim is not borne out in the results reported here. In the *Salem Witchcraft Papers*, sex has a negligible effect on relative selection. Even so, the direction of effects is the reverse of that reported by Huber (2017). This is particularly the case in non-subject relatives, where female speakers use the zero relative less frequently (and *which* more frequently) than male speakers.

3.5.3 Objective (iii): to demonstrate the value of the *Salem Witchcraft Papers* as a resource for investigating linguistic variation in 17<sup>th</sup> century Massachusetts English (and more broadly, early colonial American English) in its socio-historical context

The use of the *Salem Witchcraft Papers* as a dataset offers marked advantages in the study of restrictive relativisation. While this is a more general research objective, some discussion will be devoted to the specific advantages that this resource can offer in the examination of this variable.

As mentioned in Section 3.2.2, relativisation is not particularly frequent in running discourse, and so a large amount of data is required to obtain enough tokens for reliable quantitative analysis. The surviving documents in the *Salem Witchcraft Papers* are sufficiently lengthy to enable quantitative and statistical analysis of relativisation, refining previous claims and offering new insights into the distribution and conditioning of this well-studied variable.

The legal context of the documents has also acted to ensure the survival of a lot of sociobiographical details about the speakers, allowing the investigation of the effect of numerous social factor groups on relativiser choice. This is particularly important in establishing age effects on the use of relativisers, which are a strong indication of change in progress. Additionally, the diversity of text types, from more speech-like (examinations, depositions) to less speech-like (letters, official documents) enables the researcher to



attempt to address the incursion of the *wh*-relatives into the spoken language, a major feature of the relative system in Early Modern English.

However, as mentioned in Section 2.2, the *Salem Witchcraft Papers* should be approached with caution. The ever-present issues of ‘bad data’ and scribal interference notwithstanding, errors in the transcription of the documents may result in potentially misleading token counts. Careful attention should be paid to the fidelity of the dataset to the original manuscript, facilitated in this case through comparison with the *Records of the Salem Witch-Hunt*.

The discussion above highlights the utility of the *Salem Witchcraft Papers* to the study of relativisation, demonstrating that even with very well-studied variables, careful application of data and variationist methodology can leverage new insights.

## CHAPTER 4: *DO*-SUPPORT

### 4.1 Introduction to *Do*-Support

This chapter investigates the use of *do*-support in the *Salem Witchcraft Papers*. *Do*-support (also referred to in the literature as *do*-periphrasis, periphrastic *do*, and auxiliary *do*), is one of the most extensively studied linguistic variables in English (Bækken, 2002: 317; Moessner, 2017: 179). The patterning of *do*-support in English is far from monolithic, however, and its use as a marker of tense and aspect is reported to be sensitive to inter-varietal differences (Kortmann, 2004).

Cross-linguistic analogues of *do*-support are rare, but this phenomenon is a notable feature of Present Day English (Denison, 1993: 255; Warner, 2017: 224), with apparent cognates in all other West Germanic languages (except Afrikaans), albeit with differences in syntactic distribution (Hoekstra, 2016; Langer, 2001; Weber, 2018). In West Germanic dialects, a light verb (e.g. High German *tun*, Low German *doon*, Dutch *doen*) acts as an auxiliary, selecting a bare infinitive, much like English *do*-support (Hoekstra, 2016: 318; Langer, 2001: 12-15).

*Do*-support in the English language can be described as a construction whereby a semantically empty operator *do* is inserted above the verb (Oudesluijs et al., 2022: 4). As an operator, *do* exhibits NICE-properties (Huddleston, 1976: 333). For example, it acts as the operator in negated or inverted contexts, post-verbal ellipsis, or in emphatic contexts, and is inserted in the absence of other possible operators such as BE and the modals (Denison, 1993: 255; Los, 2015: 91-93).

- |      |                                                                        |                  |
|------|------------------------------------------------------------------------|------------------|
| (62) | a) You <b>did</b> not lock your door.<br>b) *You locked not your door. | <i>Negation</i>  |
| (63) | a) <b>Did</b> you lock your door?<br>b) *Locked you your door?         | <i>Inversion</i> |
| (64) | He asked me to lock the door and so I <b>did</b> .                     | <i>Ellipsis</i>  |

- (65) a) You should have locked the door.  
 b) But I **did** lock the door.

*Emphasis*

Negation and inversion were previously available contexts for lexical verbs (and continue to be in Dutch and German), but in English, these grammatical environments have been restricted to auxiliaries (including *do*) since the loss of V-to-I movement during the transition to Early Modern English (Los, 2015: 92, 192). *Do*-support is properly distinguished from the lexical verb *do*, which does not exhibit the same properties, and can even require *do*-support in its own right (e.g. John *didn't do* his work) (Denison, 1993: 256).

## 4.2 Review of the Literature on *Do*-Support

*Do*-support has received attention from many sub-fields of linguistics, including syntax (Akahane, 1998; Alcázar & Saltarelli, 2013; Ard, 1982; Bruening, 2010; Eitelmann, 2016; Hein, 2017; Schütze, 2004; Yamada, 2019; Ziegeler, 2004), semantics (Culicover, 2008; Hewson, 1990), and language acquisition (Al-Maktrah et al., 2017; Anderssen & Bentzen, 2018; Bahns, 1991; Pfenninger, 2014; Santelmann et al., 2002; Snyder, 2017). Other work has explored cross-linguistic counterparts of *do*-support, including in Basque (Haddican, 2007), West Germanic dialects (Hoekstra, 2016; Langer, 2001; Weber, 2018), and Northern Italian dialects (Benincà & Poletto, 2004; Swinburne, 2025).

Pertinent to the diachronic study of *do*-support in the present study is the notable attention that this phenomenon has elicited in (historical) sociolinguistics (Bækken, 2002; Cuesta, 2016; Díaz-Vera, 1999; Ihalainen, 1991; Jones & Tagliamonte, 2004; Kallel, 2002; Kroch, 1989; Nurmi, 2018; Ogura, 1993; Oudesluijs et al., 2022; Pérez, 1997; Rissanen, 1985, 2003; Söderlund, 2017; Warner, 2005, 2017; Wischer, 2008). As Söderlund (2017: 252-253) notes, much early historical research on the variable focused on literary sources, including poetry, plays and fiction. This is particularly the case for Ellegård's (1953) seminal and much-cited study of *do*-support, which forms the basis of much subsequent work on the variable. More recent work has recognised the importance of examining the historical use of *do*-support in speech-related genres like trial data (Rissanen, 2003), witness depositions (Rissanen, 1985; Söderlund, 2017), and letters (Nurmi, 2018).

As traditional historical approaches tend to focus on literary texts, the use of speech-related genres has the opportunity to uncover facets of historical language use that have been obscured by the attention paid to conventional written genres (Van der Wal & Rutten, 2013: 1-2). The use of speech-related genres (particularly trial data, which is typically recorded by a third party), also offers a unique window on to the speech of the middle and lower levels of society, who are under-represented in the historical (literary) record (Van der Wal & Rutten, 2013: 1). Additionally, if Rissanen's (1985) claim that *do*-support has its origins in spoken language is true, it is methodologically imperative to access and use historical surrogates of orality in any examination of this variable.

Despite 200 years of scholarly interest, diachronic investigations of American English constitute a fertile area for further research (Kytö & Siebers, 2022: 1-2). This is particularly the case for the early colonial period (dating back to around 1600). The *Salem Witchcraft Papers* is an as yet under-utilised resource that contains many previously unpublished documents (Archer, 2002; Díaz-Vera, 1999; Doty, 2007; Doty & Hiltunen, 2002; Grund, 2007, 2012; Hiltunen, 1996, 2010; Hiltunen & Peikola, 2007; Kahlas-Tarkka, 2012; Kytö, 2022; Peikola, 2012; Rissanen, 1985, 2003). The target phenomenon was deliberately chosen for its variable status at the time of the *Salem Witch Trials*. As a window on diachronic orality at the tail-end of the 17th century, the *Salem Witchcraft Papers* are an indispensable resource for examining the changes in this variable system.

While this variable has received much attention in the literature since Ellegård's (1953) seminal work (e.g. Bækken, 2002; Cuesta, 2016; de Both, 2019; Jones & Tagliamonte, 2004; Kallel, 2002; Kroch, 1989; Rissanen, 1985; Söderlund, 2017; Warner, 2005; Wischer, 2008), the application of variationist methodologies to the study of *do*-support in the late 17th century has not been a staple component of earlier research addressing this topic. Embedding the research in a variationist framework will enable a more fine-grained reconstruction not just of the variable system, but of the linguistic and social factors influencing the selection of *do*-support. The ensuing section details the methodological steps that were taken to achieve a better understanding of the contextual constraints operating on this variable system.

#### 4.2.1 *Do*-Support in Present Day English

The status and distribution of *do*-support in Standard Present Day English has been established since the 19<sup>th</sup> century (Jones & Tagliamonte, 2004: 96; Oudesluijs et al., 2022: 3). It is obligatory in negative (including negative imperatives) and interrogative contexts, but in affirmative declaratives, it is now restricted to emphatic use (Han & Kroch, 2000: 311-312; Partridge, 1948: 27). The latter restriction is one of the main distinctions between the use of *do*-support in Present Day and Early Modern English, where non-emphatic affirmative declarative *do* was possible (though still relatively infrequent) in literary English (Los, 2015: 110). Similarly, affirmative imperatives only allow *do*-support in emphatic contexts (Han & Kroch, 2000: 314).

Emphatic affirmative declarative *do* fulfils three functions in Present Day English. It can mark a polarity contrast (contradicting a preceding negative), indicate a tense/modality contrast (used contrastively with another auxiliary), or give exclamatory emphasis (similar to positive imperative *do*) (Denison, 1993: 266; Quirk et al., 1985). The ability to provide emphasis is a core trait of *do* and other auxiliaries, and has been attested for *do*-support as early as the fifteenth century (Denison, 1993: 266-267; Los, 2015: 93, 110). The examples below are taken from Denison (1993: 266).

- |      |                                                    |                           |
|------|----------------------------------------------------|---------------------------|
| (66) | You said I didn't, but I <i>did</i> lock the door. | (polarity contrast)       |
| (67) | I <i>did</i> and <i>do</i> take great care of it.  | (tense/modality contrast) |
| (68) | You <i>do</i> make a fuss about things.            | (exclamatory emphasis)    |

In Present Day English, *do*-support is subject to cross-varietal variation, particularly regarding the use of non-emphatic affirmative declarative *do*. In its putative place of origin in the southwest of England (de Both, 2019: 2; Ellegård, 1953: 47), *do*-support is said to have survived in a wider range of syntactic contexts than in Standard English, particularly with regard to affirmative declarative *do* (Denison, 1993: 264). For example, Ihalainen (1991: 148) notes that non-emphatic affirmative declarative *do* is still possible in present-day dialect of East Somerset. However, its use is restricted to habitual contexts, whereas the simple verb alternative is available in both habitual and perfective contexts (Ihalainen, 1991:

159). Jones and Tagliamonte (2004) also find a preference for *do*-support in habitual contexts in Present Day Somerset English. De Both (2019: 28-29) examined the use of *do*-support in present tense affirmative declaratives in the southwest of England. The majority of speakers in De Both's (2019) data use either *do* or the non-standard verbal *-s* (e.g. *I do go to the pub/I goes to the pub*). Those speakers who use both, tended to restrict the use of *do*-support to habitual contexts. Rickford (1986) suggests that habitual *be* in African American Vernacular English (AAVE) may have its origins in the construction *does be*, which is derived from the use of habitual *do*-support by Southern Irish and Southwestern English colonial speakers during the 17<sup>th</sup>-19<sup>th</sup> centuries.

Using normalised frequencies per 10,000 words, Schulz (2021: 504-505) claims that affirmative declarative *do* is less frequent in American English than it is in British English. This difference is particularly pronounced in speech, and less so in writing. American and Canadian English are claimed to show lower rates of affirmative declarative *do* than British English, while New Zealand English shows higher rates (Schulz, 2021). However, as Schulz (2021: 510) acknowledges, affirmative declarative *do* appears to be highly sensitive to genre differences, and therefore apparent differences in the use of *do*-support across dialects may be the effect of uneven distribution of text types in the ICE database. Hundt (2015) finds that in 19<sup>th</sup> century New Zealand and Australian English, *do*-support developed along an almost parallel path with British English.

Examining the use of affirmative declarative *do* in Present Day Asian Englishes, Schulz (2021) finds that the frequency and distribution of *do*-support in each variety is derived from its input variety. Therefore, Hong Kong English patterns like British English, while Philippines English patterns more like American English. Affirmative declarative *do* is generally found to be less frequent in Asian Englishes than in British English. Inter-varietal differences in the distribution of *do*-support are not restricted to affirmative declaratives; for example, Deuber (2014) finds that in both Jamaican and Trinidad English, *do*-support is absent from questions.

The findings above demonstrate that *do*-support is sensitive to cross-varietal variation, and any quantitative investigation of the variable must consider the effects of regional distribution, particularly with regard to closely-related varieties.

The use of *do*-support in modern varieties of English has also been used to infer and trace historical connections between otherwise seemingly unrelated varieties. For example, Jones and Tagliamonte (2004) compare the use of *did* in Somerset English and Samaná English (the latter spoken in the eponymous peninsula in the Dominican Republic), two varieties that have retained the earlier non-emphatic affirmative declarative *do*. These two varieties provide crucial links in establishing transatlantic historical connections between British and American English (Jones & Tagliamonte, 2004: 96).

Jones & Tagliamonte (2004: 111) found that rates of *did* in Samaná and Somerset English were almost identical, despite notable geographic and historical separation. In both data sets, *did* almost exclusively appeared in habitual contexts (Jones & Tagliamonte, 2004: 108). In fact, the patterning and distribution of *did* was remarkably similar in both varieties, being favoured in subordinate clauses, with pre-verbal adverbs, and with transitive verbs (Jones & Tagliamonte, 2004: 113-114).

The above paragraphs serve to illustrate that *do*-support in Present Day English exhibits cross-varietal idiosyncrasies. Ensuing sections will delve into claims on the origins of *do*-support (4.2.2), before sketching out the status of *do*-support chronologically, tracing the history of *do*-support from its origins to the end of the Early Modern English period (4.2.3). Discussion will then turn to some of the linguistic and social factors that have been claimed to influence the use of *do*-support in Early Modern English (4.2.4, 4.2.5), in keeping with the focus in this dissertation on late seventeenth-century American English.

## 4.2.2 The Origins of Do-Support

### 4.2.2.1 Internally-Motivated Claims

The genesis of *do*-support has received substantial attention in the literature (Bækken, 2002: 317). Several theories of origin have been proposed, and the debate is far from resolved. Ellegård (1953: 146) associates the original use of *do*-support with poetic convention, in connection with poetic metre, as does Royster (1915: 190). Other work has suggested that *do*-support reflects a language-internal evolution common to Germanic languages (Tieken-Boon van Ostade, 1987).

Ellegård (1953) proposes most persuasively that *do*-support has its origin in causative *do*, analogous to *faire* in French. After *do*-support becomes an option, it competes with causative *do* for a period. Eventually, *do*-support is successful and begins to gain ground rapidly, as causative *do* disappears from the language (Kroch, 1989: 141; Los, 2015: 110). During the period of competition between the two uses, there is evidence of forms that are ambiguous, and might be read as either causative *do* or *do*-support (e.g. example 69) (Warner, 2005; Nurmi, 1999). *Do*-support is reported to be a reanalysis of causative *do* as a semantically empty auxiliary (Denison, 1993: 279-281; Warner, 2017: 228).

- (69) but as sone as y cam to Exeter then was y yn heven: and be cause that y am now in joy i *do*  
*send* you this letter

(Annys Wydeslade – *The Stonor Letters and Papers*, 1480)

Royster (1915: 194-195) claims that the low use of *do*-support in the north of England in the fifteenth century provides evidence for a causative origin. Causative *do* was also rare in this area, which was instead dominated by the causative *gar*.

- (70) And seyne **gert** *brek* doune the vall  
*And soon did break down the wall*

(John Barbour – *The Bruce*, 1300)

Royster (1915: 195) notes that the wherever *gar* is more common (and therefore causative *do* is rare), *do*-support is also not found. Partridge (1948: 26) also indicates that *do*-support probably developed from causative *do* in Middle English.

Denison (1993: 279) goes a step further, suggesting that several early uses of *do* (causative, substitute, and periphrastic) are the same construction – *do* + infinitive. Denison (1993: 279-280) claims that these constructions all appeared almost simultaneously, and that some undefined underlying structural change occurred which led to the development of *do*-support proper in Early Modern English.



While the first attested example of *do*-support in English comes from poetry, there have been numerous claims, as alluded to above, that *do*-support has its origin as a feature of the spoken language (Rissanen, 1985: 177; Samuels, 1972: 175). Such claims cannot be substantiated with concrete evidence, as contemporaneous speech is inaccessible to the modern researcher. Rissanen (1985: 177) makes the claim, following Visser (1969: 1498), that *do*-support was most likely prevalent in the spoken language before it was first attested in writing. Samuels (1972: 175) describes *do*-support as most likely originating in the spoken language. Partridge (1948: 26) claims that avoidance of *do*-support in early written texts may suggest that the feature was quite common in the spoken language, presumably with a colloquial association, but again, the requisite evidence is lacking.

#### 4.2.2.2 *Externally-Motivated Claims*

There have been numerous attempts to connect the rise of *do*-support in English with language contact situations. One school of thought hypothesizes that English *do*-support is the result of contact with Celtic languages (Poussa, 1990; Preusler, 1956).

Preusler (1956: 334-336) argues that *do*-support entered English as a calque on an Old Welsh periphrastic construction *gwneuthur* ‘do’ + verbal noun construction. Geographical proximity with the south-west of England is claimed by Preusler (1956: 334-336) to be evidence of Welsh (and Cornish) influence on English. Ellegård (1953) argues against a Celtic origin, stating that Celtic influence would have played a minor role at best, as the earliest attestation of *do*-support in English is several hundred years after the most likely period of intense English-Celtic contact. Van der Auwera and Genée (2002: 298-299) note the possibility of such an origin, but highlight that there is no evidence to prove or disprove such a claim (see below in this section for further discussion on the methodological prerequisites to substantiate claims of language contact-induced change).

Poussa (1990) also claims a Celtic origin for *do*-support, with a much earlier origin than is otherwise attested, putting forward the notion that *do*-support was used colloquially in OE as a result of Saxon-Celtic contact. This claim is made on the basis that Somerset English is particularly distinct from other varieties of English as a result of language contact with Celtic, based on the names for bodies of water in the region (Poussa, 1990: 420). Poussa (1990) suggests that the southwest of England would have had the most prolonged period of contact with Celtic languages, and that this is the reason for *do*-support first appearing

in this area. The lack of *do*-support in Old English texts is attributed to the fact that such texts were typically literary, and were therefore less likely to use colloquial features, and that texts representing speech are very rare (Denison, 1993: 281-282).<sup>13</sup>

Similarly, Klemola (2018) suggests that Celtic influence likely contributed to the development of *do*-support as a result of the phenomenon's origins in the southwest of England. This claim, however, is based on data from the Survey of English Dialects (SED), which is both contemporary (from the middle of the 20<sup>th</sup> century) and, as Wagner (2007: 263) notes, insufficient methodologically for quantitative research.

McWhorter (2009) also argues in favour of the Celtic origin hypothesis, noting that *do*-support in Welsh and Cornish cannot be borrowed from English, as it is also present in Breton, whose speakers were removed from contact with English some 800 years before *do*-support was first attested in English. This argument does not account for the notion that *do*-support may be a separate innovation in the Celtic languages.

Van der Auwera and Genée (2002: 288-291) summarise the Celtic-origin hypotheses for the origin of *do*-support, and provide examples of relevant constructions in Celtic languages. The Celtic equivalent of *do*-support appears with a verbal noun, rather than the infinitive as in English. The *do*-verb may precede the verbal noun (attested from Old Irish onwards), or follow the verbal noun (with an intervening relative particle), which is more commonly associated with Brythonic languages (Welsh, Breton, and Cornish). Surviving examples of a Celtic equivalent of *do*-support appear to be predominantly from formal, literary texts, including prose narratives and drama. Owen (2018: 109) notes that the Middle Welsh equivalent of *do*-support is relatively rare in extant data. Van der Auwera and Genée (2002: 288) cite examples from Middle Welsh (11<sup>th</sup> century), reproduced below.

- (71)       ath           gyrchu a   wna  
          and.your   attack PT will.do  
          ‘and he will attack you’

---

<sup>13</sup> However, as Timofeeva (2022: 52) notes, the survival of legal documents (charters) and epistolary fragments does provide some limited insight into Old English orality.

- (72) kyuodi a oruc a dyuot a Lynn Cuch  
 arise PT he.did and come to Glyn Cuch  
 ‘he arose and went to Glyn Cuch’

Based on these examples, it is not clear that the English *do*-support and its Celtic equivalent are truly analogous. And, as Owen (2018: 107) demonstrates, there are notable differences between its use in, for example, Middle Welsh and Modern Welsh. Modern Celtic, then, is not suitable as evidence of *do*-support in the medieval period. Additionally, Brythonic influence on English is claimed to have occurred before the 11<sup>th</sup> century. This predates not only the appearance of *do*-support in English, but also the appearance of the equivalent construction in early Middle Welsh (Owen, 2018: 111).

While on the surface these hypotheses of contact-induced change are attractive propositions, such claims often lack rigorous empirical evidence that is necessary to demonstrate that contact is at the root of change (Poplack & Levey, 2010; Poplack et al., 2012). Poplack et al. (2012: 222-223) outline several exacting tests for empirically demonstrating that language contact has resulted in change, including:

- I. Identify a source for the change, and delineate the variable context and factors conditioning selection of variants
- II. Comparing the variable constraints in the source and host languages
- III. Comparing the variable constraints with other variants of the same construction in the host language system
- IV. Comparison with pre- and non-contact varieties of the host language to prove that the proposed change was not present
- V. Comparison with pre-contact source language to prove that the feature in question was present
- VI. Prove that the change is not a result of internal motivations (or language universals)

Overwhelmingly, the studies claiming a Celtic language-contact origin for *do*-support are not able to satisfy *any* of these claims, let alone all of them. Many of the claims are based on superficial similarities and ignore the possibility that *different* underlying grammatical processes can give rise to the *same* surface

string (Poplack 2018:28). Proponents of the Celtic origin for *do*-support often cite as evidence the fact that non-emphatic affirmative declarative *do* has survived in southwestern English (an area claimed to have had the most contact with Celtic). However, Present Day southwestern English is separated from the earliest attested appearance of *do*-support by nearly 800 years – a notably lengthy period to claim such a connection. It is evidently much easier to assert a Celtic motivation for the origins of *do*-support in the English language on the basis of superficial pattern matching rather than to systematically demonstrate that this was the case using an accountable methodology, as outlined in Poplack et al. (2012).

Having summarised some of the possible theories of origin that have been proposed for *do*-support in English, discussion will now focus on the chronological development of *do*-support over time.

#### 4.2.3 History of *Do*-Support

The chronological development of *do*-support has been the focus of much research since Ellegård's (1953) seminal work (for example, Kroch, 1989; Ogura, 1993). Much previous work has attempted to identify certain factors, both linguistic and extra-linguistic, which condition the selection of *do*-support in relevant contexts, in order to properly elucidate the extent of the variability and to characterize trajectories of linguistic change in the use of this feature. One major issue that has been addressed by historical researchers is the decline (and ultimate demise) of non-emphatic *do* in affirmative declarative contexts in Standard English, as illustrated in examples 73-75 below.

(73) She **did** Sign y<sup>e</sup> book & y<sup>e</sup> Mark was Red

(SWP019.1 – Sarah Bridges,  
Examination: 25<sup>th</sup> August, 1692)

(74) 5 afflicted persons **doe** charge this woman to be the very w <wo>man that hurts them

(SWP013.2 – Ezekiel Cheever,  
Examination: 19<sup>th</sup> April, 1692)

(75) abgeall Williams **did** one this har testimony [Lost] oath which she had taken

(SWP094.17 – author unknown,

It is clear that change has occurred in the scope of affirmative declarative *do*, and evidence of that change in progress can be found in Early Modern English data sources, both literary and speech-related.

#### 4.2.3.1 Do-Support in Old English

In Old English, there were several possible uses for *do*, including as a full verb (much like Present Day English), substitute *do* (where *do* replaces a full lexical verb to avoid repetition), and as a causative (Fischer, 1992: 267-268). Examples below are reproduced from Fischer (1992: 267-268).

- |      |                                                                                                                                            |                     |
|------|--------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| (76) | Uton... don hyne on pone ealdan pytt<br>‘Let us do [= put] him in this old well’                                                           | <i>Lexical verb</i> |
| (77) | ... he miccle ma on his dea ðe acwealde, ðonne he ær cucu dyd<br>‘... he killed many more in death than he did before [when he was] alive’ | <i>Substitute</i>   |
| (78) | And treowa he deð færllice blowan and eft raðe asearian<br>‘And trees he does [= causes] to bloom suddenly and again to wither quickly’    | <i>Causative</i>    |

Each of the above uses of *do* continued into Middle English. While competing theories on the origin of *do*-support exist (as discussed in Section 4.2.2), many linguists believe that *do*-support developed out of one of the earlier uses of *do*-support in 76-78 above (e.g. Denison, 1993; Ellegård, 1953; Traugott, 1972; Visser, 1963-73). The appearance of *do*-support in English, followed by the subsequent disappearance of causative *do*, suggests that these forms may have been in competition, and lends some credence to the notion that *do*-support is derived from causative *do*.

#### 4.2.3.2 Do-Support in Middle English

The earliest attested instance of unambiguous *do*-support is during the Middle English period, in 13<sup>th</sup> century poetry (Denison, 1993; Ellegård, 1953; Royster, 1915). Much early evidence on *do*-support comes from poetry, literature, and drama, bearing in mind the paucity of extant speech-like data surviving

from the Middle English period (with certain notable exceptions like the Paston letters, the Cely letters, and the Stonor letters; see e.g., Bergs, 2005). Royster (1915: 189) cites an early use of affirmative declarative *do* in the works of the monk and poet, John Lydgate, with an estimated date of 1387.

- (79) And as I **did** approche, Again þe sonne that shone, me þouȝt, so clere  
‘And as I did approach, again the sun that shone, seemed to me, so clear’  
(Lydgate, 1387: Line 20)

The first recorded use of *do*-support in a negative declarative is at the end of the fourteenth century, with *do*-support appearing in interrogatives at around the same time (Denison, 1993: 265). Examples of each construction in Middle English are given below (from Denison, 1993: 264-265).

*Affirmative Declarative*

- (80) His sclauyn he dude dun legge  
his pilgrim’s-cloak he did down lay  
‘He laid down his pilgrim’s cloak’  
(Horn, c1300: line 1057)

*Negative Declarative*

- (81) ... þoo persones ... þe wheche were nought present atte seide dyner ne neyther  
deden nought heren þe same acountes  
‘... those persons... who were not present at the said dinner did not hear the same accounts’  
(Book of London English,  
1423: 181.1286)

*Interrogative*

- (82) Why ... doe men put their sons to the Civill Law ... ?  
(Wimbledon Sermon,  
1635: A8r.23)

Geographically, *do*-support is first attested in the southwest of England, before spreading outwards (Royster, 1915). The distribution of *do*-support in later periods supports this origin as the phenomenon is less frequent in regions that are geographically more distant from the southwest (Ellegård, 1953). For example, Royster (1915: 193) finds frequent use of *do*-support in the works of fifteenth century writers from the Midlands. Contrastively, *do*-support is progressively rarer further to the north, and entirely absent from Middle Scots in the first half of the fifteenth century (Royster, 1915: 193; Söderlund, 2017: 273).

#### 4.2.3.3 *Do-Support in Early Modern English*

Any discussion of the development of *do*-support is incomplete without mention of Ellegård's (1953) seminal study (Figure 1 below), which draws on written data from 1390 to 1700, and whose account has been widely accepted in the literature (Moessner, 2017: 179). Ellegård's (1953) work is notable for its use of large amounts of data, and the application of quantitative methodology (though estimating rates of *do*-support based on a sample of each text means that Ellegård's approach does not respect the Principle of Accountability). It should be noted, however, that Ellegård (1953) only drew data from prose texts, and so the data reflects only the status of *do*-support in the conventional literary language of the time. The ensuing discussion will begin with Ellegård (1953) and other studies based on canonical literary texts, before considering treatments of speech-related genres such as trial proceedings and personal documents (e.g. letters and diaries). As will be shown, data from speech-related genres demonstrates different social and linguistic embedding in the use of *do*-support when compared with results from canonical written sources.

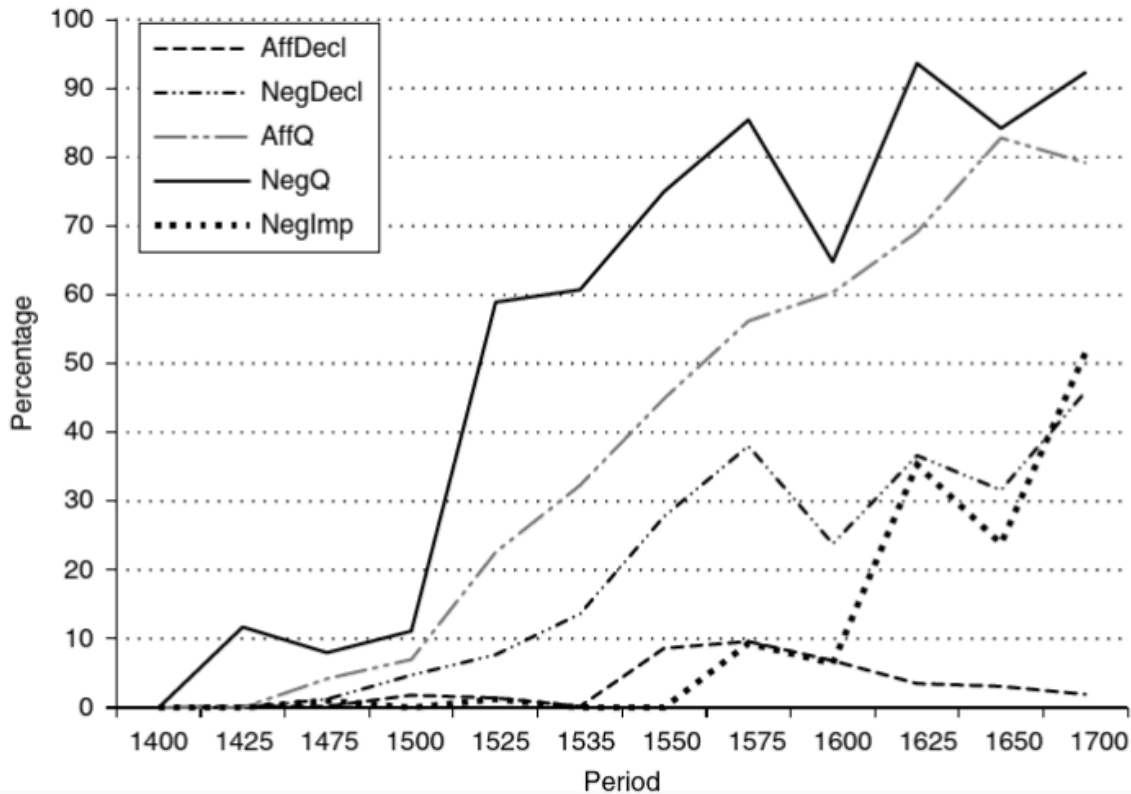


Figure 1: The distribution of *do*-support from 1400-1700 (from Los, 2015: 111, based on Ellegård, 1953).

In all contexts (interrogatives, negatives, and affirmative declaratives), *do*-support increases in frequency up until 1550-1575 (Ellegård, 1953: 166). After this point there is a split: while rates of *do*-support continue to rise in affirmative interrogatives, they begin to decline steadily in affirmative declaratives from a peak of 9.3% (Ellegård, 1953). The use of *do*-support in negative declaratives drops at the end of the 16<sup>th</sup> century (1575-1600), and does not regain its former frequency until Ellegård's (1953) final period (1650-1700). Negative interrogatives show a drop in frequency in 1575-1600, before resuming their climb upwards. In negative imperatives, use remains low (roughly commensurate with the rate of affirmative declaratives) throughout, until 1600, where there is a sudden massive increase in the use of *do*-support in negative imperatives, bringing usage in line with that of negative declaratives (Ellegård, 1953: 162).

One of the biggest questions in the development of *do*-support, then, is how to describe the behaviour and distribution of *do* in the middle of the 16<sup>th</sup> century, as it seems clear that this period witnessed key changes. Drawing on the data used by Ellegård (1953), Kroch (1989: 154-156) claims that the use of *do*-support in all environments is introduced by a single grammatical rule, meaning that it is available in all



environments initially, not appearing sequentially as has previously been suggested (e.g. Bailey, 1973: 82). Kroch (1989) describes this as the Constant Rate Effect, whereby an ongoing change maintains the same pattern of favouring and disfavouring effects across contexts. Where processing constraints affect one environment (i.e. by favouring the use of *do*-support), they affect the frequency in all environments simultaneously, with changes spreading at the same rate across contexts. Under Kroch's (1989) model, rates of change proceed linearly regardless of context. To explain the behaviour of affirmative declaratives after 1560, Kroch (1989) proposes that grammatical reorganization takes place, leading to different grammatical rules governing the use of *do*-support in declaratives and interrogatives. This allows rates of declarative *do* to decline separately from interrogatives. An additional rule would then be applied to separate affirmative declaratives from negative declaratives, allowing the latter to continue to appear more frequently with *do*-support.

Ogura (1993: 63) challenges Kroch's approach, arguing instead that the change did in fact proceed sequentially. Utilising Ellegård's (1953) data, Ogura (1993: 67-68) finds "a clear correlation between the starting points and the rates of change" – earlier changes proceed more slowly, and later changes change more rapidly. *Do*-support is reliably attested in affirmative declaratives first, followed by negative declaratives, then negative interrogatives, affirmative questions, and finally negative imperatives. Ogura (1993: 68-70) finds that the later changes show the steepest slope, increasing at the greatest rate. This convincingly explains the very sudden increase in Ellegård's (1953) data in the use of negative imperatives after 1600. Ogura (1993: 67) attributes the decline in *do*-support after 1550 to competition between this form and the simple form. After a brief decline, *do*-support steadily increases in negatives and interrogatives. The cause for the steady decline in affirmative declarative *do* is not clear. Kroch (1989) claims a reanalysis of the English auxiliary system during the 16<sup>th</sup> century, resulted in affirmative declarative *do* coming to be treated as a separate construction by the grammar.

Contrastively, Warner (2005: 259-260; 2017) points to a sociolinguistic change rather than a grammatical one, proposing that *do*-support was subject to a shift in register. Before the mid-16<sup>th</sup> century, *do*-support was associated with informal contexts, consistent with Rissanen's (1985) claim that *do*-support originated in the spoken language (Nurmi, 1999: 106). In the middle of the 16<sup>th</sup> century, affirmative declarative *do* appears to receive an increased association with more formal, literary contexts, and begins to decline in frequency in informal contexts (Warner, 2005; 2017). However, it should be noted that this claim is based

on Ellegård's (1953) database, where informal and speech-like genres are conspicuously absent. As a result, based on work by Biber (1988: 108), Warner (2005: 260) operationalises lexical complexity as a measure of formality, under the assumption that high informational density is a hallmark of more formal textual styles.

It was noted previously that Ellegård (1953), and subsequently Kroch (1989) and Ogura (1993), based their claims on the distribution of *do*-support on explicitly written texts and prose. More recent work, drawing on more speech-related text types, as opposed to canonical literary sources, has found that the peak for affirmative declarative *do* may be somewhat later than has traditionally been posited, and also that this usage persists for slightly longer than conventionally claimed. For example, Nurmi (1999), drawing on data from the *Corpus of Early English Correspondence* and the *Helsinki Corpus*, dates the decline of *do*-support in affirmative declaratives to the early seventeenth century, after a peak in the final decade of the sixteenth century, claiming that the frequency drop is connected to political change. As Nurmi (1999) observes, the timing of the drop coincides with the accession of James VI and I, uniting the crowns of England and Scotland. Nurmi (1999) claims that this introduces a prestige variety of English spoken by the Scottish nobility with lower rates of *do*-support. It does not seem clear, however, that this would have any widespread effect on actual usage, beyond a putative impact on a small group of peers, and it therefore seems an unlikely catalyst for such a major change for all users of the language.

Drawing on data on affirmative declaratives in the *Corpus of Early English Correspondence Extension* (covering the period from 1681 to 1800), Nurmi (2018: 125-126) finds that at the end of the seventeenth century rates of *do*-support in that syntactic context are almost equivalent to its peak in affirmative declaratives, while the eighteenth century displays a rapid decline, approaching Present Day English rates in the latter half of the eighteenth century.

Oudesluijs et al. (2022) examine the use of *do*-support in declaratives in fifteenth to eighteenth century civic records and letters from the civic centres of Bristol, Coventry, and York. In Bristol, rates are generally low, with negative declarative *do* being rarer than affirmative declarative *do* (Oudesluijs et al., 2022: 12-13). There is a decline in the use of affirmative declarative *do* in the seventeenth century. In Coventry as well, *do*-support predominates in affirmative declaratives, which may be reflective of its use in formulaic, set phrases, such as *do covenant and grant*, *do appear*, and *do set and let farm* (Oudesluijs

et al., 2022: 15). In the York data *do*-support is generally rarer, which is in line with claims that *do*-support is less frequent in northern varieties of English (Oudesluijs et al., 2022: 17).

Oudesluijs et al. (2022) note that the use of *do*-support in trials, laws and diaries is relatively stable during the period when affirmative declarative *do* is declining in most other text types (1550-1700). There are at least two possible explanations for this. One is that legal genres (e.g. trials and laws) may be more conservative in the retention of affirmative declarative *do* (Romaine, 1982: 97; Tiersma, 2015). Alternatively, given that trial data is often used to approximate spoken language, and personal texts like diaries and letters are thought to fairly closely represent speech (Schneider, 2002), this could be suggestive of a change happening in the written language. According to this hypothesis, affirmative declarative *do* would persist longer in more speech-like texts. These ideas are not exclusive, and both may be at work during this period. The latter idea in particular might be supported by the fact that the peak (and subsequent decline) reported for affirmative declaratives is earlier in writing than speech-related texts by 50-100 years, but this is at odds with the generally more conservative rate of change found in writing (Nurmi, 1999: 128).

Söderlund (2017) examines the use of *do*-support in witness depositions from *An Electronic Text Edition of Depositions 1560-1760* (ETED), which contains depositions from London and the North, South, and East of England (Kytö et al., 2007: 71-74). In this speech-related data, affirmative declarative *do* only begins to decline towards the end of the 17<sup>th</sup> century, supporting the hypothesis that this decline happened later in the spoken language. In negative declaratives, *do*-support increases steadily, reaching a rate of 75.4% from 1700-1760, a period which is not covered by Ellegård's data. Söderlund (2017: 263-264) also examined the use of *do*-support with the *know* verbs in negative declaratives (a group of verbs including *know*, *make*, *come*, and main verb *do*), which have previously been claimed to appear less frequently with *do* (Ellegård, 1953; Nurmi, 1999). In this speech-related data, the previous claims are initially borne out, with the *know*-group showing lower rates of *do*-support. While the *know*-group may have initially been resistant to *do*-support, this resistance appears to have been nullified during the 17<sup>th</sup> century, as by the 18<sup>th</sup> century (1700-1760 in Söderlund's data), rates of *do*-support with the *know*-group are in line with other verbs at around 75%. Interrogatives were relatively rare in Söderlund's (2017) deposition data, but do show an increase in the use of *do*-support throughout the period. It should be noted that due to the relative

rarity of interrogatives (which is attributed to genre effects), and to enable comparison with Ellegård (1953), Söderlund (2017) did not examine negative and affirmative interrogatives separately.

Summarizing, it can be seen, then, that the Early Modern English period is one of notable change in the use of *do*-support, particularly across sentence types (declarative, interrogative, imperative). The general trend for the period is a steep decline in the use of affirmative declarative *do*, and an increase in the rate of *do*-support in negatives and interrogatives (Ellegård, 1953). As will be demonstrated, these shifts are not restricted to British English, but are apparent in colonial varieties of English as well.

Of paramount importance to the present study is the research carried out by Rissanen (1985; 2003), which is centred on early American colonial English. Using records of church meetings in 16<sup>th</sup> century New England and contemporary history books, Rissanen (1985) finds a higher rate for affirmative declarative *do* than that reported by Ellegård (1953) (17% compared with Ellegård's peak of 10%). Rissanen (1985: 169) connects this high rate to the nature of the data, where solemn declarations (frequent in the church meeting records) appear to favour the use of *do*-support. Rissanen (1985: 177) also affirms the relation between the use of *do*-support in affirmative declaratives and the spoken language.

Rissanen (2003) looks at several linguistic features, including *do*-support, in an earlier, non-electronic version of the *Salem Witchcraft Papers*. The study is not intended to be an exhaustive account of *do*-support in the *Salem Witchcraft Papers* so much as an illustration of the types of insight on Earlier American English that those historical documents could offer. As such, Rissanen (2003) offers an initial strategic foundation for a more detailed examination of *do*-support in the *Salem Witchcraft Papers*, as undertaken in the present dissertation.

Rissanen (2003) uncovers several interesting findings which warrant further investigation. For example, Rissanen (2003: 108) finds that *do*-support is most common with interrogatives. In Rissanen's data, 98% of questions appear with *do*-support. It should be noted, however, that Rissanen (2003: 108) only counted interrogatives marked by a question mark.<sup>14</sup> *Do*-support is also found to be frequent with negatives,

---

<sup>14</sup> In the *Salem Witchcraft Papers* many questions are not marked with punctuation. While the Early Modern English period is associated with the regularisation of punctuation (and spelling), 17<sup>th</sup> century texts have been noted as being somewhat

appearing in 80% of instances (Rissanen, 2003: 107-108). A majority (70%) of those negatives that appear in the simple form (i.e. without *do*-support), occur with the main verb *know* – which, as discussed earlier, is reported to resist the use of *do*-support (Ellegård, 1953; Nurmi, 1999).

Another claim made by Rissanen (2003: 108) is that 17<sup>th</sup> century American English uses more affirmative declarative *do* than equivalent British data. This is based on British data in an earlier paper (Rissanen, 1991), where there are <30 instances of *do*-support in affirmative declaratives per 10,000 words. By comparison, in the data from Salem, Rissanen (2003: 108) reports 51 occurrences per 10,000 words, which Rissanen takes as evidence that American English is lagging behind British English in the decline (and restriction to emphatic contexts) of affirmative declarative *do*. Normalised frequencies are not a methodologically reliable approach for exploring linguistic variables, as they do not account for the instances where the variant of choice could have been used but was not, a violation of the Principle of Accountability (Labov, 1972). Rissanen (2003: 108-109) does suggest that the higher rates of affirmative declarative *do* in the *Salem Witchcraft Papers* might be a consequence of the conventions of legal language, citing its use in high-frequency phrases such as *did torture*, *did afflict*, and *did own*.

Summarising, it seems clear that there is a need for further work on the use of *do*-support during the key period of Early Modern English. In particular, there is scope for examination of speech-related genres, which appear to demonstrate different patterns of *do*-support when compared with canonical written texts.

Additionally, much previous work on *do*-support has not been carried out within a variationist sociolinguistic framework, and so it is hampered methodologically by non-adherence to the Principle of Accountability (Labov, 1972), which requires the inclusion of all instances where the linguistic variable under investigation could have been used (even if it was not). For example, Nurmi (1999) calculates rates of *do*-support using normalised frequencies per 10,000 words. This means that the instances where *do*-support could have been used but was not, are omitted from the data, impeding comparability across studies and potentially confounding the results of such comparisons.

---

inconsistent in their use of punctuation (Bernal, 2016; Salmon, 1988; Sánchez, 2024; Thaisen, 2020). This will likely result in Rissanen (2003) under-estimating the number of interrogatives in the data.

Furthermore, many studies on *do*-support focus on its use in one particular syntactic context – most often in affirmative declaratives (e.g. Bækken, 2002; Ihalainen, 1991; Nurmi, 1999, 2018; Rissanen, 1985; Wischer, 2008), but also sometimes negatives (Díaz-Vera, 1999; Pérez, 1997; Warner, 2005). This focus does not contextualise the findings within the wider envelope of variability and limits the possibilities for systematic comparisons to be drawn across disparate studies. An exhaustive account of the use of *do*-support in each of the relevant syntactic contexts in which it appears in Early Modern English speech-related data, in keeping with the Principle of Accountability (Labov, 1972), has the potential to leverage new insights into the development of *do*-support.

#### 4.2.4 Internal Factors Affecting the Selection of *Do*-Support in Early Modern English

From the extant literature, it is possible to abstract several factors that have been claimed to influence the use of *do*-support. These include internal (linguistic) and external (social) factors. Chief among these, as has already been discussed, is the use of *do*-support in different syntactic contexts, which has been well documented in the literature. Negatives and interrogatives might be expected to show far higher rates of *do*-support than affirmative declaratives, which should exhibit very low rates of *do*-support, based on a review of the historical literature. In the ensuing sections, potential predictors contributing to the selection of *do*-support are reviewed.

##### 4.2.4.1 *Verb Morphology*

Among the linguistic factors that are implicated in *do*-support, many involve characteristics of the main verb. For example, Bækken (2002: 326) finds that morphologically irregular verbs favour the use of *do*-support. This is attributed to the increasing importance of *do*-support for tense-marking in these contexts. Ellegård (1953: 180) also found that *do*-support was preferred with verbs that do not exhibit morphological differences between their past and present tense forms (e.g. *put*, *cut*, *set*). It seems likely that this behaviour is connected. These facts are important in the context of wider changes in the English language. As Denison (1993: 283) notes, this coincides with “a period of great uncertainty as to tense marking on irregular and Latinate verbs”. The use of *do*-support as a bearer of tense-marking lends stability to the process by helping to disambiguate tense distinctions that might otherwise remain ambiguous.

#### 4.2.4.2 Tense

The association of *do*-support with temporal reference is foregrounded in the literature. *Do*-support is repeatedly reported to be favoured in the past tense (Bækken, 2002: 325-326). Wischer (2008: 146) finds that irregular verb forms favour *do*-support particularly in past tense contexts. Söderlund (2017: 261) observes that where *do*-support is used in affirmative declaratives, this is primarily in the past tense. Söderlund (2017) connects this finding with the fact that depositions are mainly in the past tense (as they are recounting past events). Contrastively, Oudesluijs et al. (2022: 17) find that in their York data, *do*-support is quite common in affirmative declaratives in the present tense. Wischer (2008: 146) also finds affirmative declarative *do* to be slightly more frequent in the present tense. These discrepant findings merit further investigation. The current study puts these contradictory claims to the test.

#### 4.2.4.3 Aspect

Closely related to tense is the association of *do*-support with habitual aspect. Jones and Tagliamonte (2004: 108-109), who only looked at the past tense, found that *do*-support “overwhelmingly encoded habitual aspect”. By contrast, Wischer (2008: 147-148) finds that *do*-support expresses habitual aspect in the present tense, and perfective aspect in the past tense. The findings from these two studies are clearly at odds. It should be noted that Wischer (2008) examined Early Modern English data, while Jones and Tagliamonte (2004) used modern data from Samaná and Somerset. A possible explanation for the disparity in these findings, then, could be that previously (i.e. in Early Modern English), *do*-support encoded perfective aspect in the past tense, but underwent a shift to represent the habitual in both the present and past tenses. Samuels (1972: 176) notes that 17<sup>th</sup> century grammarians describe *do*-support as having an imperfective (i.e. habitual) reading. The association of *do*-support with the habitual, at least in the 17<sup>th</sup> century, was therefore sufficiently salient to have been noticed by grammarians.

#### 4.2.4.4 Adverbs

The co-occurrence of *do*-support with pre-verbal adverbs (intervening between *do* and the verb) in the Early Modern English period is well-documented (Rissanen, 1985: 164-165). This has been connected with the later emphatic use of *do*-support (Stein, 1990: 358). Bækken (2002: 328) and Nurmi (2018: 132) find that affirmative declaratives conspicuously take *do*-support more frequently when there is an adverb, particularly when they are intensifying or ‘emotional’ adverbs, as the addition of an adverb lengthens the verb phrase, ‘adding weight’. Bækken (2002: 327-328) does not specify the criteria for an

adverb to be considered emotional, but the examples that they give all contain adverbs of manner (e.g. *mournfully, more bitterly, steadfastly*). Both Rissanen (1985: 173) and Söderlund (2017: 269) find the presence of pre-verbal adverbs to correlate with an increase in the rate of *do*-support in affirmative declaratives, with the latter noting a growth in the rate of adverbs with affirmative declarative *do* during the period 1560-1760. Partridge (1948: 27) notes a connection between standard non-emphatic affirmative declarative *do* and certain conjunctive adverbs beyond the eighteenth century. In interrogatives with an adverb, *do*-support is found to be more frequent by Ellegård (1953: 205).

#### 4.2.4.5 Clause Structure

The interaction of *do*-support with clause structure has been the locus of conflicting claims in the literature. *Do*-support has been separately claimed to be associated with main clauses (Stein, 1990), and with indicating a return to a main clause after a subordinate clause (Samuels, 1972), or with coordinated clauses (Denison, 1993). Drawing on quantitative data from Early American English, Rissanen (1985) finds that *do*-support is favoured in subordinate clauses, and particularly so in relative clauses. Ihalainen (1991) also looked at Early Modern English data, and found that *do*-support frequently occurs in subordinate clauses. Jones and Tagliamonte (2004: 113) find that *do*-support is favoured in subordinate clauses in Samaná and Somerset English. This effect is reported by those authors to be statistically significant.

#### 4.2.4.6 Lexical Effects

The use of *do*-support has been associated with specific lexical collocations. Collocational patterns are difficult (but not impossible) to approach from a quantitative perspective, but as Jones and Tagliamonte (2004: 102) note, the idiosyncrasy of verb collocations with *do*-support makes lexical patterning an important part of the linguistic conditioning of the variable.

As noted above, a limited group of verbs are said to have resisted *do*-support, which Ellegård (1953) referred to as the *know*-group. In negative declaratives this group includes the verbs *know, do, care, doubt, fear, list, skill, boot, come*, and *speak*. Similarly, for interrogatives *come, like, make*, and *say* are claimed to be less frequent with *do*-support. The *know*-group is thought to lag behind other verbs in the use of *do*-support by about 200 years, and this lag has been linked to the high frequency nature of some of these verbs (Ogura, 1993: 76). Pérez (1997: 40) finds that verbs including *have, need, doubt, know, dare*, and



*mean* rarely appear with *do*-support in negative declaratives. Nurmi (2018: 122) finds that in affirmative declaratives, verbs expressing mental processes (*want*, *think*, *seem*, *say*, and most interestingly, *know*) appear more frequently with *do*-support. This is obviously in contrast to claims made about the relationship between *know* and *do*-support, and may reflect the effect of sentence type. *Know* may be resistant to *do*-support in negatives and interrogatives, but favour the use of *do*-support in declaratives. This potential disconnect warrants further investigation.

Other verbs seem to favour the selection of *do*-support. As mentioned, Nurmi (2018: 122) claims that verbs expressing mental processes favour *do*-support. Ellegård (1953: 172) finds that *do*-support often co-occurs with verbs like *beseech*, *believe*, *confess*, and *preach*. Warner (2017: 230-231) notes the frequency of *do*-support with *preach*, *eat*, and *think*.

As Warner (2017: 231) indicates, however, it may not be the verb itself that resists *do*-support, but its use in a particular construction (e.g., with negative polarity, as suggested by the preservation of the simple form (albeit archaically) in fixed phrases, including *know not*, *care not*, *doubt not*, and *mistake not* (Partridge, 1948: 27).

#### 4.2.5 External Factors Affecting the Selection of *Do*-Support in Early Modern English

Numerous external or social factors have been argued to influence the use of *do*-support in Early Modern English. These include geographical region, sex, formality, and age. The impact of social factors may be diachronically constrained. Nurmi (2018: 126) finds the rates of *do*-support in affirmative declaratives to be essentially unaffected by social embedding in the eighteenth century, suggesting that late-stage changes lose their social indexation. If that is so, then data from the sixteenth and seventeenth centuries could provide crucial insights into the social embedding of *do*-support in those periods.

##### 4.2.5.1 *Geographical Origin*

As observed earlier, Ellegård (1953) claimed that *do*-support was more common in the south of England, and less frequent in northern varieties of English (Söderlund 2017: 273). This is in keeping with the received wisdom that *do*-support would be less frequent in those regions which are more

geographically distant from its purported origins in the south-west. Oudesluijs et al. (2022: 20-21), however, did not find region to be a reliable indicator of the use of *do*-support.

#### 4.2.5.2 Age

The age of a speaker has been connected with the use of *do*-support, particularly in Early Modern English, when this feature was undergoing change. Kallel (2002: 177), using epistolary data from the 16<sup>th</sup> century, shows that younger speakers use *do*-support significantly more frequently than older speakers; this is particularly the case with negatives. Kallel (2002) notes the interplay between age and sex, as younger females use *do*-support disproportionately more than their male contemporaries. Warner (2005: 268) observes that from 1600 onwards, age becomes a highly significant predictor of *do*-support use, with younger generations successively employing *do*-support with increasing frequency.

#### 4.2.5.3 Sex

Nurmi (1999: 154, 171) finds that, in declaratives, *do*-support is used far more frequently by female speakers after 1620. This is in stark contrast to earlier usage, where male speakers are claimed to use more *do*-support (Nurmi, 1999). The reasons for such a drastic switch are unclear. Nurmi (2018: 129) states that female speakers appeared to be responsible for the rise in the use of *do*-support in the 16<sup>th</sup> and 17<sup>th</sup> centuries – however, by the end of the 17<sup>th</sup> century and into the 18<sup>th</sup> century, sex is reported to have ceased to be a significant influence on the use of *do*-support (Nurmi, 2018). The *Salem Witchcraft Papers*, dating from 1692-1693, are uniquely positioned to examine this claim in further detail.

#### 4.2.5.4 Document Type

*Do*-support is reported to have originated as a colloquial form (Ellegard, 1953), as mentioned earlier. Nurmi (2018: 127) notes a preference for *do*-support in 18<sup>th</sup> century letters sent to more intimate personal connections (family members and close friends) compared to letters to more distant acquaintances, but this difference is not statistically significant. Nurmi (1999: 106) finds affirmative declarative *do* to be more frequent in informal letters in the sixteenth century. However, this relationship is not stable throughout the period. For example, Warner (2017: 230) observes that *do*-support appears to accrue prestige in the second half of the sixteenth century, as determined by its increase in frequency in formal letters during this period. Similarly, Cuesta (2016: 133) connects the use of *do*-support in religious language in the north of England in the sixteenth century with formality. Wischer (2008: 142-143) claims

that *do*-support is more frequent in formal contexts in the early sixteenth century, and becomes increasingly associated with informal settings up until the early eighteenth century.

One possible diagnostic of formality lies in the use of *do*-support in different text types (Nurmi, 2000). Pérez (1997) looks at the use of *do*-support in negative declaratives in the Early Modern English (1500-1710) part of the *Helsinki Corpus of English Texts*, comprising a number of different text types. *Do*-support was found to be more frequent in texts that more closely approximate speech (trials), less frequent in formal text types like handbooks and educational documents, and was rare in text types based on fictional speech (fiction and drama).

Text type also has an effect on the frequency of conditions (such as sentence type) which are highly propitious to the use of *do*-support. For example, Rissanen (1991) finds that trial data has an elevated frequency of negative constructions. Negative constructions are known to favour the selection of *do*-support, and so it might be expected that these texts would demonstrate higher rates of *do*-support than fiction or drama, which Rissanen (1991) finds to have lower rates of negative clauses. As Söderlund (2017: 252-253) highlights, much of the previous work on *do*-support has focused on either written genres, or text types featuring constructed speech (like fiction, poetry, and plays). Following Culpeper and Kytö (2010), Söderlund (2017: 275) demonstrates that witness depositions pattern more closely with constructed speech genres like fiction than with more authentic speech-related genres (like trial proceedings). The differential behaviour of these texts with regard to the selection of *do*-support opens up additional avenues of inquiry.

#### 4.2.6 The *Salem Witchcraft Papers* as a Window on the Development of *Do*-Support

It has been demonstrated that there are gaps in the literature on *do*-support, most notably with regard to the application of quantitative analyses to Early Modern English speech-related data, and to Early American English, which have been neglected (see Kytö & Siebers, 2022).

The *Salem Witchcraft Papers* provides a large amount of data representing different text types, spanning the formal/informal continuum, and including both speech-related (e.g. examinations and depositions) and more conventional written genres (e.g. warrants, indictments). The data come from the end of the 17<sup>th</sup>

century, the final time period examined in Ellegård's (1953) analysis, representing an era where affirmative declarative *do* is thought to have been in steep decline, while negative and interrogative *do* are claimed to be increasingly frequent. This period has the potential to leverage new insights into the variable use of *do*-support during the later stages of change.

The *Salem Witchcraft Papers* offer unique data on early American English, providing the best opportunity to reconstruct 17<sup>th</sup> century spoken norms in Massachusetts English (Rissanen, 2003: 88). As noted previously (Sections 4.2.4 and 4.2.5), many of the factors that have been claimed to condition the selection of *do*-support have generated inconsistent—and at times contradictory—results in the literature. The large amount of varied data that the *Salem Witchcraft Papers* provide, allows for a more robust investigation into the variable structure of *do*-support in the 17<sup>th</sup> century and re-examination of some of the conflicting claims regarding its use.

The current research will build on the preliminary investigation on *do*-support in the *Salem Witchcraft Papers* carried out by Rissanen (2003). This earlier investigation was based on an incomplete version of the *Salem Witchcraft Papers*, and so features only a limited dataset. The current research will provide a more comprehensive analysis, addressing the available data in the digital version of the *Salem Witchcraft Papers*.

This chapter will build on earlier findings, while also extending the investigation to other under-studied constraints reported to influence the selection of *do*-support. The application of multivariate analysis to quantitative data will offer new insights into the specific factor groups conditioning variant choice, leveraging empirically-informed findings to achieve a clearer understanding of *do*-support in Earlier American English.

### **4.3 Methodology**

#### **4.3.1 Variable Context**

With the intended goal of reconstructing the variable system influencing the use of *do*-support in 17<sup>th</sup> century English, all eligible tokens have been extracted from the *Salem Witchcraft Papers*. *Do*-support is here defined as the use of an auxiliary *do* in affirmative and negative declaratives, interrogatives, and

imperatives. In accordance with the Principle of Accountability (Labov, 1972: 72), eligible tokens also include the simple form (i.e. those contexts where *do*-support could have been used, but was not). Pertinent examples from the data illustrating the targeted environments are given below.

*Affirmative Declarative*

- (83) a. now this man **did** Say before he went; y<sup>t</sup> he would fetch blood of her  
(SWP013.10 – Samuel Shattuck,  
Testimony: 2<sup>nd</sup> June, 1692)
- b. Cap<sup>t</sup> Putman testified that he made his wife enter into a Covenant  
(SWP022.4 – author unknown,  
Examination: 9<sup>th</sup> May, 1692)

*Affirmative Interrogative*

- (84) a. what **doe** you say to these murders you are charged with  
(SWP013.2 – John Hawthorne,  
Examination: 19<sup>th</sup> April, 1692)
- b. what evil spirit have you familiarity with  
(SWP063.2 – John Hawthorne,  
Examination: 1<sup>st</sup> March, 1692)

*Affirmative Imperative*

- (85) a. **doe** tell us then  
(SWP063.2 – John Hawthorne,  
Examination: 1<sup>st</sup> March, 1692)
- b. Now tell us how we shall know Who doth hurt these if you do not  
(SWP038.2 – John Hawthorne,  
Examination: 21<sup>st</sup> March, 1692)

*Negative Declarative*

(86) a. I never **did** hurt them in my life I **did** never see these persons before  
(SWP013.2 – Bridget Bishop,  
Examination: 19<sup>th</sup> April, 1692)

b. I know not whether to make any returns.  
(SWP026.1 – Dudley Broadstreet,  
Examination: 10<sup>th</sup> August, 1692)

*Negative Interrogative*

(87) a. oh goode Bishop **did** you not come to our house the Last night  
(SWP013.2 – Mercy Lewis,  
Examination: 19<sup>th</sup> April, 1692)

b. Have not you compassion for these afflicted  
(SWP092.3 – author unknown,  
Examination: 2<sup>nd</sup> May, 1692)

*Negative Imperative*

(88) a. Yes Mother **doe** not deny it  
(SWP025.2 – Mary Lacey Jr.,  
Examination: 22<sup>nd</sup> July, 1692)

b. Charge him not unless it be he.  
(SWP005.1 – author unknown,  
Examination: 22<sup>nd</sup> April, 1692)

All tokens were manually extracted from the online digital version of the *Salem Witchcraft Papers*. A second reading was undertaken as part of the quality control procedures implemented to check the validity of extracted tokens and their eligibility for inclusion in the study. Several instances of *do* (and the simple form) were removed from the data as they were either ambiguous or outside of the envelope of variability (for example, matrix verb *do*). These exclusions are illustrated below:

### *Matrix Verb Do*

- (89) And shall **do** and receive that which by the said Court shall then & there be Enjoyed him  
(SWP006.4 – John Richards,  
Recognizance: 31<sup>st</sup> December, 1692)

### *Do As Part of a Light Verb Construction*

- (90) Did you go & **do hurt** with Sarah Good?  
(SWP069.1 – John Hawthorne,  
Examination: 19<sup>th</sup> April, 1692)

### *Repetition*

- (91) Sarah ~~Good~~ Osburne vpon Examination denied y<sup>e</sup> matter of fact (viz) y<sup>t</sup> she ever  
vnderstood or vsed any Witchcraft **or vsed any Witchcraft**  
(SWP095.1 – John Hawthorne,  
Examination: 1<sup>st</sup> March, 1692)

## 4.3.2 Internal Factor Groups

### *4.3.2.1 Sentence Type*

The 13,000 tokens that were retained for analysis were coded for several factor groups abstracted from the extant literature on *do*-support. Chief among these factor groups, and present in almost every piece of research on the variable, is the primordial effect of sentence type on the selection of *do*-support. As indicated earlier, the status of *do*-support in the 17<sup>th</sup> century is characterized by a marked decline in the use of *do* in affirmative declaratives (Ellegård, 1953; Kroch, 1989; Söderlund, 2017). At the same time, the *use* of *do*-support in negatives and interrogatives is reported to be on the rise (Kroch, 1989; Rissanen, 1999: 245; Söderlund, 2017: 253). All tokens are coded for clause type (declarative, interrogative, and imperative) and clause polarity (affirmative or negative).

### *4.3.2.2 Clause Type*

Despite the prodigious amount of attention lavished on the influence of clause structure on *do*-support, this remains an area where there are multiple inconsistencies in the literature. Somewhat

contradictorily, *do*-support has been associated with main clauses (Stein, 1990: 32-34), as well as subordinate clauses (Ihalainen, 1991; Jones & Tagliamonte, 2004; Rissanen, 1985; Samuels, 1972). Rissanen (1985) finds the putative subordinate clause effect to be particularly strong with relative clauses. It should be noted that Stein's (1990) claim is based on data from narratives, and therefore may not be fully representative of what obtains in more speech-like sources. In order to explore these discrepant findings further, tokens were coded for whether the clause containing *do*-support is a main clause, a relative clause, or another type of subordinate clause.

#### *Main Clause*

- (92) What **do** you say of those murders you are charged with?

(SWP013.1 – John Hawthorne,  
Examination: 19<sup>th</sup> April, 1692)

#### *Relative Clause*

- (93) She hes seen no appearance since but a ffly which **did** speake to her

(SWP008.1 – Mary Barker,  
Examination: 29<sup>th</sup> August, 1692)

#### *Other Subordinate Clauses*

- (94) How is it then, that your appearance **doth** hurt these?

(SWP013.1 – John Hawthorne,  
Examination: 19<sup>th</sup> April, 1692)

#### *4.3.2.3 Main Verb Morphology*

Tokens were coded according to key morphological distinctions in the main verb. The distinction is between weak verbs (those that take the suffix *-ed* in the past tense, such as *talk*, *torment*, *greet*), strong verbs (those that form the past tense via vowel gradation or ablaut), such as *eat*, *sing*, *speak*), and semi-weak verbs (including *keep/kept* and *weep/wept*). An additional category was included as there were several instances of verbs that have strong verbal morphology in Standard Present Day English, but were used with weak verbal morphology in the data. Such verbs can still be treated as weak in non-standard varieties, and include examples like *catched* (*caught*), *drawed* (*drew*), and *telled* (*told*).



The consideration of main verb morphology builds on the claims by Bækken (2002: 326) and Ellegård (1953: 180) that *do*-support is favoured with morphologically strong verbs. The prediction is that *do*-support will be more frequent in those contexts as support for tense-marking (Denison, 1993: 283).

#### *Weak Verbal Morphology*

- (95) the Devil brought my Shape to Salem, and did **afflict** M. S. and R. F.

(SWP009.4 – William Barker Sr.,  
Confession: 29<sup>th</sup> August, 1692)

#### *Strong Verbal Morphology*

- (96) I never did **hurt** them in my life I did never see these persons before

(SWP013.2 – Bridget Bishop,  
Examination: 19<sup>th</sup> April, 1692)

#### *Semi-Weak Verbal Morphology*

- (97) and then I p<sup><f>^</sup>{u}rsued it: and it **Kept** y<sup>e</sup> same Distance in y<sup>e</sup> parth before me

(SWP032.5 – Thomas Boarman,  
Deposition: 4<sup>th</sup> April, 1692)

#### *Strong verb in Present Day English treated as weak in 17<sup>th</sup> century American English*<sup>15</sup>

- (98) it stood y<sup>r</sup> in pots & they **drawed** it out of a barrill

(SWP018.1 – Mary Bridges jr.,  
Examination: 25<sup>th</sup> August, 1692)

---

<sup>15</sup> Past tense formation in Early Modern English admitted competition between strong and weak forms for many verbs (Cheshire, 1994: 116).

#### 4.3.2.4 Tense

Tense has been connected with the use of *do*-support in the literature. In particular, *do*-support has been associated with the past tense (Bækken, 2002: 325-326; Söderlund, 2017: 261). Wischer (2008: 146) connects the use of *do*-support in past tense contexts with its use with irregular verb forms, finding that these factor groups are complementary. Oudesluijs et al. (2022: 17) and Wischer (2008: 146) both find affirmative declarative *do* to be more frequent in present tense contexts. Accordingly, to explore the potential effects of temporal reference on variant selection, tokens were coded for present and past tense.<sup>16</sup>

#### *Present Tense*

(99) If I **do** speak you will not beleive me?

(SWP024.3 – Martha Carrier,  
Examination: 31<sup>st</sup> May, 1692)

#### *Past Tense*

(100) What **did** you think of the actions of others before your sisters came out, **did** you think it was Witchcraft?

(SWP045.2 – John Hawthorne,  
Examination: 22<sup>nd</sup> April, 1692)

#### 4.3.2.5 Aspect

In a similar vein, *do*-support is closely associated with habitual aspect (Jones & Tagliamonte, 2004: 108-109; Wischer, 2008: 147-148). However, Wischer (2008) finds that *do*-support is often used to express habitual aspect in the present tense, while Jones and Tagliamonte (2004) claim that *do*-support encodes habitual aspect in the past tense (but do not themselves include present-tense contexts). To explore these claims further, tokens were coded according to whether they occurred in habitual or non-habitual contexts.

#### *Habitual*

---

<sup>16</sup> It should be noted that future tense contexts are not part of the variable context as the presence of auxiliaries *will* and *shall* blocks the use of *do*-support.

(101) and I knowing of her well also y<sup>e</sup> garb she **did** vse to goe in

(SWP013.13 – Richard Coman,  
Testimony: 2<sup>nd</sup> June, 1692)

#### *Non-Habitual*

(102) on y<sup>e</sup> after noon of s<sup>d</sup> day the dores being shutt I **did** see a black pig in the Roome

(SWP013.11 – John Louder,  
Testimony: 2<sup>nd</sup> June, 1692)

#### *4.3.2.6 Adverbs*

There is a long-standing and well-documented relationship between *do*-support and the presence of adverbs (Rissanen, 1985: 164-165). All tokens were coded for the presence (or absence) of an adverb modifying the main verb, under the hypothesis that *do*-support will be favoured with adverbs, particularly those that are intensifying (Ellegård, 1953: 205). This could demonstrate a connection between *do*-support and its later emphatic use. It is difficult to objectively separate emphatic and non-emphatic uses of *do*-support without access to audio recordings, and so the effect of adverbs is operationalised as a measure of emphasis. If this is the case, the effect of adverbs on the use of *do*-support might be expected to be stronger with affirmative declaratives (i.e. the context where *do*-support would later become entirely restricted to emphatic contexts in Standard English) (Bækken, 2002: 328; Nurmi, 2018: 122; Stein, 1990: 358). However, both Rissanen (1985: 173) and Söderlund (2017: 269) found adverbs to be a non-statistically significant predictor of the selection of *do*-support.

#### *Adverb*

(103) This Deponet **doth** *veryly* beleive that the said Bridget Bishop was Instumentall to his  
Daughter Prisillas Death

(SWP013.8 – William Stacey,  
Deposition: 30<sup>th</sup> May, 1692)

#### *No Adverb*

(104) I **doe** tell the truth I never hurt these persons

(SWP013.2 – Bridget Bishop,  
Examination: 19<sup>th</sup> April, 1692)

#### 4.3.2.7 Adverb Position

Also considered was adverb position relative to the main verb. Pre-verbal adverbs have previously been associated with the use of *do*-support (Ihalainen, 1976; Pérez, 1997; Wischer, 2008). Tokens were coded as either pre-verbal, post-verbal, or pre- and post-verbal, which encompasses cases such as (107), where there are multiple adverbs in a sentence, appearing simultaneously in pre-verbal and post-verbal positions.

##### *Pre-Verbal*

(105) and I Afirmed to her face to be tru and y<sup>t</sup> I **did plainly** see her

(SWP013.11 – John Louder,  
Testimony: 2<sup>nd</sup> June, 1692)

##### *Post-Verbal*

(106) and then he sett upon me most grevively and **did** tortor me **most dreadfully**

(SWP107.12 – Ann Putnam jr.,  
Deposition: 11<sup>th</sup> April, 1692)

##### *Pre- & Post-Verbal*

(107) I saw the Apperishtion of gooddy procktor amongst the wictes ~~which~~ {& she} **did almost**  
choake me **Immediatly**

(SWP106.14 – Ann Putnam jr.,  
Deposition: 11<sup>th</sup> April, 1692)

#### 4.3.3 Lexical Collocations

*Do*-support has also been connected with numerous lexical collocations. Such lexical patterning is reported to be an important facet of variable usage (Jones & Tagliamonte, 2004: 102). Certain verbs (e.g.

the *know*-group described by Ellegård, 1953) have been claimed to resist *do*-support, lagging behind other verbs in the use of *do*-support by 200 years (Ogura, 1993: 76). Interestingly, Nurmi (2018: 122) finds that *know*, among other verbs, favours the use of *do*-support in affirmative declaratives. While this is considerably more difficult to quantify than other factor groups (due to the sheer number of different verbs that *do*-support appears with), each main verb in the data was recorded, so that the relationship between *do*-support and each verb could be examined quantitatively. High-frequency verbs (i.e. those appearing more than 100 times in the data, which includes the *know*-group) will be examined in more detail, including cross-tabulation with sentence type. The effect of language of origin (Germanic or Latinate) on *do*-support will also be investigated, as this has been claimed to influence the use of *do*-support (Samuels, 1972).

#### 4.3.4 External Factor Groups

The data were also coded for extra-linguistic (i.e. social) factors, many of which have been reported to influence the use of *do*-support. Where socio-biographical information is inferable from the data (or exists elsewhere in the historical record), such information is recorded for each speaker. In many cases (e.g. for birthplace and age), not enough information exists to categorize every speaker reliably. In such instances, the pertinent speaker information is coded as ambiguous.

##### 4.3.4.1 *Speaker Sex*

Male and female speakers were coded separately. Speaker sex has been reported to influence the selection of *do*-support in Early Modern English, with female speakers using more *do*-support than male speakers (Nurmi, 2018: 129), in line with the oft-reported tendency of female speakers to spearhead linguistic change (Labov, 1990). The effect of speaker sex on *do*-support is reported to have ceased to be operative by the 18<sup>th</sup> century (Nurmi, 2018: 129).

##### 4.3.4.2 *Speaker Age*

Speakers were classified as belonging to one of three age groups – young (0-29 years), middle-aged (30-49 years), and older (50+ years). Children are generally underrepresented in the data, and so are subsumed under the young category. The use of an apparent-time approach may shed light on inter-generational changes in the use of *do*-support. It is hypothesized, for example, that older speakers will use more affirmative declarative *do* than younger speakers. Warner (2005: 268-269) finds that in the 17<sup>th</sup>

century, younger speakers use significantly more *do*-support in negative declaratives than older speakers. Kallel (2002: 177) also notes a preference for *do*-support in negatives among younger speakers. Kallel (2002) finds that this effect is particularly pronounced in discourse associated with female speakers.

#### 4.3.4.3 *Speaker Birthplace*

The use of *do*-support has been claimed to be highly sensitive to regional differentiation (Ellegård, 1953). A number of the residents of Salem in 1692, particularly among the older generations, were not born there, having moved from Britain (e.g. Giles Corey; Francis Dane Sr; George Burroughs). In Early Modern British English, *do*-support was most common in the southwest, and was reportedly less frequent in the language of speakers from the north of the country (Ellegård, 1953; Söderlund, 2017). Speakers were coded based on their county of birth. Those who were born in New England were grouped together. A small minority of speakers (e.g. the slaves Candy and Tituba) were born elsewhere (South America and Barbados), and so those speakers have been coded as ‘other’.

#### 4.3.4.4 *Speaker Social Status*

Speakers were grouped in a reconstructed social hierarchy based on three primary indices: potential access to wealth, potential access to education, and access to specific labour skills, all of which have been associated with social capital in the region and time period (Mallinson, 2007: 150; Zemsky, 1969). Wealthier, more educated speakers were placed in the highest status group, which included government officials, members of the clergy, and members of the nobility. The middle group included the merchant class, and those belonging to a specific trade (e.g. coopers, inn owners, smiths). The lower status group included general labourers and farmhands, while the lowest status group represented the enslaved peoples of Salem.

#### 4.3.4.5 *Text Type*

The *Salem Witchcraft Papers* provide a range of genres and textual styles. These texts have been coded along a continuum representing different levels of formality and proximity to speech. Examinations represent the least formal and closest to the original speech act (Rissanen, 2003: 88). Witness depositions and testimonies are slightly more formal and further separated from the speech act (Grund et al., 2014: 44-45). Epistolary documents like letters, and petitions, bear a degree of formality (depending on the recipient), and potentially do not represent an original speech act (Nurmi, 2018). Official and legal

documents (such as indictments and warrants) lie at the most formal end of the scale, and are the least speech-like (Grund, 2021: 87). *Do*-support has been particularly associated with less formal contexts, and those that more closely approximate speech (Nurmi, 2018; Pérez, 1997; Rissanen, 1985; Wischer, 2008), though some researchers associate the feature with more formal contexts (Cuesta, 2016; Warner, 2017).

Summarizing, *do*-support is a well-studied variable, but many previous treatments of this variable are rooted in disparate frameworks of analysis, and there is consequently a distinct lack of methodological and analytical uniformity, which confounds efforts to achieve a broad consensus on the evolution of this variable. The present investigation builds on previous sociolinguistic work on *do*-support in Early Modern English, particularly Rissanen (1985; 2003) and Söderlund (2017), and seeks to challenge, corroborate, and extend previous findings.

## 4.4 Results

### 4.4.1 Overall Distribution

Table 41 shows the overall distribution of *do*-support in the *Salem Witchcraft Papers* data (set phrases have been excluded: see Section 4.4.1.1). *Do*-support accounts for 14% (N=1845) of all tokens. This rate is in line with similar results reported in the literature. For example, Söderlund (2017: 258) reports a rate of 13.6% for *do*-support for the period 1650-1699. In comparison, Ellegård (1953: 159) has a much lower rate of 6.2% *do*-support for the same period. However, it is entirely plausible that Ellegård's (1953) lower rate of *do*-support is a result of methodological differences, because, as previously observed, Ellegård (1953) examined prose texts rather than the speech-related ones targeted in the present study. It would be expected that conventional written genres would be more conservative in terms of variable behaviour implicated in change than speech-related texts. Söderlund (2017), by contrast, draws on data from witness depositions, and so his results, based as they are on speech-related sources, will likely be closer to those found in the *Salem Witchcraft Papers*.

Nurmi (1999) looks at *do*-support in the *Corpus of Early English Correspondence* (1417-1499), but broad-brush comparisons with the present investigation are impeded on several counts by the different methodological approach adopted in Nurmi (1999). Nurmi (1999: 233) reports rates for *do*-support as normalised frequencies per 10,000 words (with all the attendant issues that such an approach entails).

| <b>Variant</b>    | <b>%</b>   | <b>N</b>     |
|-------------------|------------|--------------|
| <i>do-support</i> | 14         | 1845         |
| <i>simple</i>     | 86         | 11292        |
| <b>Total</b>      | <b>100</b> | <b>13137</b> |

Table 41: Overall distribution of variants

#### 4.4.1.1. Set Phrases

*Do-support* in the data from the *Salem Witchcraft Papers* was observed to co-occur in several formulaic phrases, provided below.

| <b>Set Phrase</b>                                                 | <b><i>Do-Support</i></b> |
|-------------------------------------------------------------------|--------------------------|
| <i>NAME acknowledge(s) the truth of what they owned</i>           | 13% (5/40)               |
| <i>NAME affirm(s) the truth of what they owned</i>                | 8% (12/144)              |
| <i>... and therefore crave(s) justice</i>                         | 0% (0/17)                |
| <i>NAME declare(d) this to be true</i>                            | 7% (3/41)                |
| <i>NAME deposeth that...</i>                                      | 0% (0/5)                 |
| <i>As the law direct(s)</i>                                       | 0% (0/7)                 |
| <i>Therefore fail not...</i>                                      | 0% (0/39)                |
| <i>NAME said as follows</i>                                       | 11% (2/19)               |
| <i>NAME make(s) complaint/oath/return</i>                         | 31% (37/234)             |
| <i>NAME hereby oblige(s) themselves</i>                           | 75% (6/8)                |
| <i>If the above bound perform the above-mentioned articles...</i> | 63% (5/8)                |
| <i>The jurors present that...</i>                                 | 39% (47/122)             |
| <i>Stand(s) charged with</i>                                      | 6% (6/105)               |
| <i>NAME testifyeth that...</i>                                    | 10% (41/392)             |
| <i>The testimony of NAME witnesseth and sayeth that...</i>        | 6% (1/18)                |

Table 42: Distribution of *do-support* in set phrases

As Table 42 above demonstrates, many of the set phrases do still exhibit variation between *do-support* and the simple form, and so have been retained. Exceptions to this, which are excluded from any further analysis, are the phrases *and therefore crave(s) justice*, *NAME deposeth that...*, *as the law direct(s)*, and *therefore fail not*, which categorically take the simple form.



#### 4.4.2 Internal Factors

##### 4.4.2.1 Sentence Type and Polarity

The impact of sentence type and polarity on the use of *do*-support has been well documented in the literature (Ellegård, 1953: 161; Kroch, 1989; Ogura, 1993; Oudesluijs et al., 2022; Söderlund, 2017: 260). Discussion will focus first on sentence type, then polarity, and finally on the interplay of sentence type and polarity.

Table 43 shows that sentence type has a clear effect, with *do*-support being favoured in interrogatives (45%). Declaratives slightly favour *do*-support with a rate of 10%. *Do*-support is sparsely instantiated in imperatives (4%).

| Factors       | <i>do</i> -support |             | simple    |              | Total      |              |
|---------------|--------------------|-------------|-----------|--------------|------------|--------------|
|               | %                  | N           | %         | N            | %          | N            |
| Declarative   | 10                 | 1077        | 90        | 10125        | <b>85</b>  | <b>11202</b> |
| Interrogative | 45                 | 757         | 55        | 925          | <b>13</b>  | <b>1682</b>  |
| Imperative    | 4                  | 11          | 96        | 242          | <b>2</b>   | <b>253</b>   |
| <b>Total</b>  | <b>14</b>          | <b>1845</b> | <b>86</b> | <b>11292</b> | <b>100</b> | <b>13137</b> |

Table 43: Contribution of sentence type to the selection of *do*-support

Turning to the effects of polarity, *do*-support shows a clear preference for negative contexts, accounting for almost half of all negative tokens (46%). Affirmative contexts, by comparison, disfavour the selection of *do*-support.

| Factors      | <i>do</i> -support |             | simple    |              | Total      |              |
|--------------|--------------------|-------------|-----------|--------------|------------|--------------|
|              | %                  | N           | %         | N            | %          | N            |
| Affirmative  | 11                 | 1374        | 89        | 10734        | <b>92</b>  | <b>12108</b> |
| Negative     | 46                 | 471         | 54        | 558          | <b>8</b>   | <b>1029</b>  |
| <b>Total</b> | <b>14</b>          | <b>1845</b> | <b>86</b> | <b>11292</b> | <b>100</b> | <b>13137</b> |

Table 44: Contribution of polarity to the selection of *do*-support

Sentence type (in particular, interrogatives) and polarity (negatives) exert a strong influence on the selection of *do*-support, an effect which is amplified when the two factor groups are cross-tabulated with one another. As Table 45 demonstrates, in affirmatives contexts, *do*-support is disfavoured with declaratives (8%) and imperatives (1%), and is favoured with interrogatives (39%). In negative sentences, *do*-support is favoured regardless of sentence type. This effect is weakest in imperatives, where *do*-support accounts for 19% of tokens, and is strongest with interrogatives, where *do*-support occurs at a rate of 92%. *Do*-support is used in 37% of negative declaratives. These results suggest that negative polarity and interrogatives are two independently operating factors influencing the selection of *do*-support, which, when combined, show the most favourable conditions for the use of *do*-support.

The high rate of *do*-support with negative interrogatives dovetails with results by Ellegård (1953: 166), who also reports a rate of 92% in the period 1650-1700. Ellegård (1953: 166) gives a rate of 1.8% *do*-support in affirmative declaratives, which is markedly less than the 8% reported here. Interestingly, Ellegård (1953) finds *do*-support with affirmative interrogatives to be nearly twice as frequent (79%) as in the *Salem Witchcraft Papers* (39%).

For the equivalent time period, Söderlund (2017: 260) reports a rate of 10.3% for affirmative declarative *do*-support, and a rate of 43.2% for negative declarative *do*-support. These findings are in line with the results in Table 4. Söderlund (2017) only reports 24 total interrogative tokens for this period (all of them selecting *do*-support), a number that is too small to offer any definitive conclusions or allow for comparison.

Throughout the remaining results section, a factor will be considered to favour *do*-support if the rate of *do*-support is higher than the rate for each sentence type given in Table 45 below. Cells that are particularly pertinent to the discussion will be shaded for ease of reference.

| Factors | <i>do</i> -support |   | simple |   | <b>Total</b> |   |
|---------|--------------------|---|--------|---|--------------|---|
|         | %                  | N | %      | N | %            | N |

|                           |           |             |           |              |            |              |
|---------------------------|-----------|-------------|-----------|--------------|------------|--------------|
| Affirmative declarative   | 8         | 783         | 92        | 9625         | <b>79</b>  | <b>10408</b> |
| Affirmative interrogative | 39        | 590         | 61        | 911          | <b>11</b>  | <b>1501</b>  |
| Affirmative imperative    | 1         | 1           | 99        | 198          | <b>2</b>   | <b>199</b>   |
|                           |           |             |           |              |            |              |
| Negative declarative      | 37        | 294         | 63        | 500          | <b>6</b>   | <b>794</b>   |
| Negative interrogative    | 92        | 167         | 8         | 14           | <b>1</b>   | <b>181</b>   |
| Negative imperative       | 19        | 10          | 81        | 44           | <b>0.4</b> | <b>54</b>    |
| <b>Total</b>              | <b>14</b> | <b>1845</b> | <b>86</b> | <b>11292</b> | <b>100</b> | <b>13137</b> |

Table 45: Contribution of sentence type and polarity to the selection of *do*-support

#### 4.4.2.2 Verb Morphology

In the data considered here, both weak (15%) and strong (12%) verbs show little deviation from the overall distribution rates for *do*-support. Semi-weak verbs show a slight preference for *do*-support at 19%, but with only 10 occurrences of semi-weak verbs in the data, this effect is not robust. An additional finding of some interest is the complete absence of *do*-support in the non-standard grouping, a collection of verbs that would have strong verb morphology in standard Present Day English, but were used with weak verb morphology in the data (e.g. *caught* vs. *caught*, *telled* vs. *told*, *seed* vs *saw*) (Cheshire, 1994: 116). While 12 tokens are insufficient to make any substantive claims about this trend, this remains a topic that merits further investigation.

| Factors | <i>do</i> -support |     | simple |      | <b>Total</b> |             |
|---------|--------------------|-----|--------|------|--------------|-------------|
|         | %                  | N   | %      | N    | %            | N           |
| Weak    | 15                 | 795 | 85     | 4378 | <b>39</b>    | <b>5173</b> |

|              |           |             |           |              |            |              |
|--------------|-----------|-------------|-----------|--------------|------------|--------------|
| Strong       | 12        | 940         | 88        | 6854         | <b>59</b>  | <b>7794</b>  |
| Semi-weak    | 19        | 10          | 81        | 43           | <b>0.4</b> | <b>53</b>    |
| Non-standard | 0         | 0           | 100       | 12           | <b>0.1</b> | <b>12</b>    |
| Ambiguous    | 95        | 100         | 5         | 5            | <b>0.8</b> | <b>105</b>   |
| <b>Total</b> | <b>14</b> | <b>1845</b> | <b>86</b> | <b>11292</b> | <b>100</b> | <b>13137</b> |

Table 46: Contribution of verb strength to the selection of *do*-support

*Do*-support has previously been associated with strong verb morphology (Bækken, 2002; Ellegård, 1953). In data from 1635-1670, Bækken (2002: 326) finds that *do*-support appears with a strong or semi-weak verb 86% of the time (Bækken 2002 conflates the two, so it is not possible to tease out the respective rates). It should be noted, however, that Bækken (2002) only examined *do*-support in affirmative declaratives. Cross-tabulation of verb strength with affirmative declaratives in the *Salem Witchcraft Papers*, facilitating a closer comparison with the results obtained by Bækken (2002), shows that in this context, *do*-support is disfavoured with strong and semi-weak verbs (4% and 5% respectively), and, by comparison, is more likely to occur with weak verbs (11%). This is clearly at odds with the claims made by Bækken (2002).

| Factors      | <i>do</i> -support |            | simple    |             | Total      |                           |
|--------------|--------------------|------------|-----------|-------------|------------|---------------------------|
|              | %                  | N          | %         | N           | %          | N                         |
| Weak         | 11                 | 475        | 89        | 3840        | <b>41</b>  | <b>4315</b>               |
| Strong       | 4                  | 251        | 96        | 5734        | <b>58</b>  | <b>5985</b>               |
| Semi-Weak    | 5                  | 2          | 95        | 35          | <b>0.3</b> | <b>37</b>                 |
| Non-Standard | 0                  | 0          | 100       | 11          | <b>0.1</b> | <b>11</b>                 |
| Ambiguous    | 92                 | 55         | 8         | 5           | <b>1</b>   | <b>60</b>                 |
| <b>Total</b> | <b>8</b>           | <b>783</b> | <b>92</b> | <b>9625</b> | <b>100</b> | <b>10408<sup>17</sup></b> |

Table 47: Contribution of verb strength to the selection of *do*-support in affirmative declaratives

<sup>17</sup> The lower total number of tokens is due to the omission of interrogatives, imperatives, and negative declaratives in this table.

#### 4.4.2.3 Tense and Aspect

The *Salem Witchcraft Papers* data show a preference for *do*-support in the present tense (20%), while in past tense contexts *do*-support is comparatively disfavoured (12%). Tokens where the tense may sometimes be ambiguous (e.g. with a verb like *hurt*), while accounting for a small proportion of the data, do not appear with *do*-support.

The literature makes inconsistent claims with regard to the influence of tense on the use of *do*-support. Söderlund (2017: 261) does not fully investigate the effect of tense, but notes that affirmative declarative *do* is favoured in the past tense, accounting for 86% of tokens. Oudesluijs et al. (2022: 17), on the other hand, find that affirmative declarative *do* mainly appears in the present tense (though specific rates are missing from their results).

| Factors      | <i>do</i> -support |             | simple    |              | Total      |              |
|--------------|--------------------|-------------|-----------|--------------|------------|--------------|
|              | %                  | N           | %         | N            | %          | N            |
| Past         | 12                 | 1236        | 88        | 8821         | 77         | 10057        |
| Present      | 20                 | 609         | 80        | 2464         | 23         | 3073         |
| Ambiguous    | 0                  | 0           | 100       | 7            | 0.05       | 7            |
| <b>Total</b> | <b>14</b>          | <b>1845</b> | <b>86</b> | <b>11292</b> | <b>100</b> | <b>13137</b> |

Table 48: Contribution of tense to the selection of *do*-support

Cross-tabulating tense with sentence type affords a more fine-grained perspective on the results. Setting aside imperatives (which occur only in the present tense), it can be seen that past tense contexts favour *do*-support in affirmative interrogatives (43%), negative declaratives (34%), and especially negative interrogatives (96%). Affirmative declaratives disfavour the use of *do*-support in past tense contexts (7%). In the present tense, *do*-support is still favoured in affirmative and negative interrogatives, though the effect is less strong than with past tense contexts. *Do*-support is favoured in present tense negative declaratives (43%), at a higher rate than in the past tense (34%). Compared with the past tense, the present tense is slightly more hospitable to the selection of *do*-support in affirmative declaratives (10%).

| Factors                   | Past               |             |           |             | Present            |            |           |             | Total      |              |
|---------------------------|--------------------|-------------|-----------|-------------|--------------------|------------|-----------|-------------|------------|--------------|
|                           | <i>do</i> -support |             | simple    |             | <i>do</i> -support |            | simple    |             |            |              |
|                           | %                  | N           | %         | N           | %                  | N          | %         | N           | %          | N            |
| Affirmative Declarative   | 7                  | 600         | 93        | 8018        | 10                 | 183        | 90        | 1602        | <b>79</b>  | <b>10408</b> |
| Affirmative Interrogative | 43                 | 344         | 57        | 455         | 35                 | 246        | 65        | 456         | <b>11</b>  | <b>1501</b>  |
| Affirmative Imperative    | 0                  | 0           | 0         | 0           | 1                  | 1          | 99        | 198         | <b>2</b>   | <b>199</b>   |
|                           |                    |             |           |             |                    |            |           |             |            |              |
| Negative Declarative      | 34                 | 176         | 66        | 343         | 43                 | 118        | 57        | 155         | <b>6</b>   | <b>794</b>   |
| Negative Interrogative    | 96                 | 116         | 4         | 5           | 85                 | 51         | 15        | 9           | <b>1</b>   | <b>181</b>   |
| Negative Imperative       | 0                  | 0           | 0         | 0           | 19                 | 10         | 81        | 44          | <b>0.4</b> | <b>54</b>    |
| <b>Total</b>              | <b>9</b>           | <b>1236</b> | <b>67</b> | <b>8821</b> | <b>5</b>           | <b>609</b> | <b>19</b> | <b>2464</b> | <b>99*</b> | <b>13137</b> |

Table 49: Contribution of tense and sentence type to the selection of *do*-support

*Do*-support has been strongly associated with the habitual aspect in the past tense, most notably by Jones and Tagliamonte (2004: 108-109), who report that in their data on Somerset and Samaná English, *do*-support conveys habitual aspect in almost every instance, with the exception of only 10 tokens. Such a strong finding is clearly at odds with the results in Table 9, where habitual contexts only modestly favour the selection of *do*-support (18%) over non-habitual ones (12%). Moreover, habitual aspect is generally infrequent, and accounts for only 3% of all instances of *do*-support in the data.

| Factors  | <i>do</i> -support |    | simple |     | Total    |            |
|----------|--------------------|----|--------|-----|----------|------------|
|          | %                  | N  | %      | N   | %        | N          |
| Habitual | 18                 | 46 | 82     | 214 | <b>3</b> | <b>260</b> |

|              |    |             |    |             |            |                           |
|--------------|----|-------------|----|-------------|------------|---------------------------|
| Non-habitual | 12 | 1127        | 88 | 8459        | <b>95</b>  | <b>9586</b>               |
| Ambiguous    | 30 | 63          | 70 | 148         | <b>2</b>   | <b>211</b>                |
| <b>Total</b> | 12 | <b>1236</b> | 88 | <b>8829</b> | <b>100</b> | <b>10057<sup>18</sup></b> |

Table 50: Contribution of habitual aspect to the selection of *do*-support in past tense contexts

Breaking down the effect of habitual aspect on the use of *do*-support according to multiple syntactic environments reveals that there is a clear distinction between declaratives and interrogatives, regardless of polarity. Declaratives show higher rates of *do*-support in habitual contexts, while interrogatives show higher rates of *do*-support in non-habitual contexts. This is particularly the case with negative interrogatives.

| Factors                   | Habitual           |           |          |            | Non-Habitual       |             |           |             | Total      |                          |
|---------------------------|--------------------|-----------|----------|------------|--------------------|-------------|-----------|-------------|------------|--------------------------|
|                           | <i>do</i> -support |           | simple   |            | <i>do</i> -support |             | simple    |             |            |                          |
|                           | %                  | N         | %        | N          | %                  | N           | %         | N           | %          | N                        |
| Affirmative Declarative   | 12                 | 24        | 88       | 182        | 7                  | 561         | 93        | 7715        | <b>86</b>  | <b>8482</b>              |
| Affirmative Interrogative | 35                 | 14        | 65       | 26         | 42                 | 299         | 58        | 413         | <b>8</b>   | <b>752</b>               |
|                           |                    |           |          |            |                    |             |           |             |            |                          |
| Negative Declarative      | 40                 | 4         | 60       | 6          | 33                 | 165         | 67        | 330         | <b>5</b>   | <b>505</b>               |
| Negative Interrogative    | 80                 | 4         | 20       | 1          | 97                 | 102         | 3         | 3           | <b>1</b>   | <b>110</b>               |
| <b>Total</b>              | <b>0.4</b>         | <b>46</b> | <b>2</b> | <b>215</b> | <b>11</b>          | <b>1127</b> | <b>86</b> | <b>8461</b> | <b>100</b> | <b>9849<sup>19</sup></b> |

Table 51: Contribution of aspect and sentence type to the selection of *do*-support

<sup>18</sup> Present tense tokens are omitted from this table.

<sup>19</sup> Present tense and ambiguous tokens are omitted from this table.

#### 4.4.2.4 Clause Type

The use of *do*-support has independently been associated with main clauses (Stein, 1990), subordinate clauses (Ihalainen, 1991; Jones & Tagliamonte, 2004: 113), and relative clauses (Rissanen, 1985). Data from the *Salem Witchcraft Papers* suggests that while main clauses and subordinate clauses show little effect, relative clauses, by contrast, are actually less favourable to the use of *do*-support, at a rate of 7%.

| Factors            | <i>do</i> -support |             | simple    |              | Total      |              |
|--------------------|--------------------|-------------|-----------|--------------|------------|--------------|
|                    | %                  | N           | %         | N            | %          | N            |
| Main clause        | 15                 | 1338        | 85        | 7348         | <b>66</b>  | <b>8686</b>  |
| Subordinate clause | 13                 | 415         | 87        | 2782         | <b>24</b>  | <b>3197</b>  |
| Relative clause    | 7                  | 92          | 93        | 1162         | <b>10</b>  | <b>1254</b>  |
| <b>Total</b>       | <b>14</b>          | <b>1845</b> | <b>86</b> | <b>11292</b> | <b>100</b> | <b>13137</b> |

Table 52: Contribution of clause type to the selection of *do*-support

Cross-tabulating the effect of clause type with sentence type reveals interesting patterns that would otherwise have remained obscured. Imperatives were only found in main clauses, and so will not be discussed further in this section.

Among affirmative declaratives, *do*-support is slightly more common with other subordinate clauses (9% compared to 7% each for main and relative clauses). *Do*-support with affirmative interrogatives is favoured in main clauses (78%), but much less so in subordinate clauses (25%), and disfavoured in relative clauses (6%). Negative declaratives show only slight differentiation between clause types – 38% for main clauses, 34% for subordinate clauses, and 33% for relative clauses. Negative interrogatives are clearly very hospitable to the use of *do*-support in main and subordinate clauses. There are only two instances of a negative interrogative in a relative clause (one with *do*-support, one without), which is too few to elucidate the use of *do*-support in that context.



| Factors                   | Main               |             |           |             | Subordinate        |            |           |             | Relative           |           |          |             | Total      |              |
|---------------------------|--------------------|-------------|-----------|-------------|--------------------|------------|-----------|-------------|--------------------|-----------|----------|-------------|------------|--------------|
|                           | <i>do</i> -support |             | simple    |             | <i>do</i> -support |            | simple    |             | <i>do</i> -support |           | simple   |             |            |              |
|                           | %                  | N           | %         | N           | %                  | N          | %         | N           | %                  | N         | %        | N           | %          | N            |
| Affirmative Declarative   | 7                  | 514         | 93        | 6617        | 9                  | 230        | 91        | 2453        | 7                  | 39        | 93       | 555         | <b>79</b>  | <b>10408</b> |
| Affirmative Interrogative | 78                 | 484         | 22        | 135         | 25                 | 67         | 75        | 197         | 6                  | 39        | 94       | 579         | <b>11</b>  | <b>1501</b>  |
| Affirmative Imperative    | 1                  | 1           | 99        | 198         | 0                  | 0          | 0         | 0           | 0                  | 0         | 0        | 0           | <b>2</b>   | <b>199</b>   |
|                           |                    |             |           |             |                    |            |           |             |                    |           |          |             |            |              |
| Negative Declarative      | 38                 | 215         | 62        | 347         | 34                 | 66         | 66        | 127         | 33                 | 13        | 67       | 26          | <b>6</b>   | <b>794</b>   |
| Negative Interrogative    | 93                 | 114         | 7         | 8           | 91                 | 52         | 9         | 5           | 50                 | 1         | 50       | 1           | <b>1</b>   | <b>181</b>   |
| Negative Imperative       | 19                 | 10          | 81        | 44          | 0                  | 0          | 0         | 0           | 0                  | 0         | 0        | 0           | <b>0.4</b> | <b>54</b>    |
| <b>Total</b>              | <b>10</b>          | <b>1338</b> | <b>56</b> | <b>7349</b> | <b>3</b>           | <b>415</b> | <b>21</b> | <b>2782</b> | <b>1</b>           | <b>92</b> | <b>9</b> | <b>1161</b> | <b>100</b> | <b>13137</b> |

Table 53: Contribution of clause type and sentence type to the selection of *do*-support

#### 4.4.2.5 Adverbs

The association of *do*-support with adverbs has been connected with its later emphatic use (Stein, 1990: 358). The *Salem Witchcraft Papers* data does not show adverbs to have any appreciable effects on the selection of *do*-support. However, certain researchers (e.g. Bækken, 2002; Ellegård, 1953; Nurmi, 2018; Partridge, 1948) have suggested that the effect of adverbs on the use of *do*-support may be dependent on the type of sentence.

| Factors | <i>do</i> -support |   | simple |   | Total |   |
|---------|--------------------|---|--------|---|-------|---|
|         | %                  | N | %      | N | %     | N |

|              |           |             |           |              |            |              |
|--------------|-----------|-------------|-----------|--------------|------------|--------------|
| Adverb       | 15        | 371         | 85        | 2099         | <b>19</b>  | <b>2470</b>  |
| No adverb    | 14        | 1474        | 86        | 9193         | <b>81</b>  | <b>10667</b> |
| <b>Total</b> | <b>14</b> | <b>1845</b> | <b>86</b> | <b>11292</b> | <b>100</b> | <b>13137</b> |

Table 54: Contribution of adverbs to the selection of *do*-support

Cross-tabulation of sentence type with the presence of an adverb shows that the presence of an adverb favours the use of *do*-support in affirmative declaratives (12%). This corresponds to the findings reported by Bækken (2002: 328), Nurmi (2018: 122), and Partridge (1948: 27), who also find a correlation between affirmative declarative *do* and adverbs. Conversely, negative declaratives show a notably higher rate of *do*-support when there is no adverb (43% compared to 17% with an adverb). Affirmative interrogatives favour the use of *do*-support with an adverb (50%), while negative interrogatives show little deviation in the use of *do*-support regardless of the presence of an adverb. This broadly dovetails with Ellegård's (1953: 205) claim that *do*-support is more common in interrogatives with an adverb. In fact, the division seems to be sensitive to polarity: generally, affirmative contexts favour *do*-support with an adverb, while negative contexts disfavour. This is not an effect that appears to have been previously reported in the literature.

| Factors                   | Adverb             |     |        |      | No Adverb          |     |        |      | Total     |              |
|---------------------------|--------------------|-----|--------|------|--------------------|-----|--------|------|-----------|--------------|
|                           | <i>do</i> -support |     | simple |      | <i>do</i> -support |     | simple |      |           |              |
|                           | %                  | N   | %      | N    | %                  | N   | %      | N    | %         | N            |
| Affirmative Declarative   | 12                 | 255 | 88     | 1838 | 6                  | 523 | 94     | 7787 | <b>79</b> | <b>10408</b> |
| Affirmative Interrogative | 50                 | 58  | 50     | 58   | 38                 | 531 | 62     | 853  | <b>11</b> | <b>1501</b>  |
| Affirmative Imperative    | 3                  | 1   | 97     | 32   | 0                  | 0   | 100    | 166  | <b>2</b>  | <b>199</b>   |
|                           |                    |     |        |      |                    |     |        |      |           |              |
| Negative Declarative      | 17                 | 29  | 83     | 146  | 43                 | 265 | 57     | 354  | <b>6</b>  | <b>794</b>   |

|                        |          |            |           |             |           |             |           |             |            |              |
|------------------------|----------|------------|-----------|-------------|-----------|-------------|-----------|-------------|------------|--------------|
| Negative Interrogative | 86       | 19         | 14        | 3           | 93        | 148         | 7         | 11          | <b>1</b>   | <b>181</b>   |
| Negative Imperative    | 0        | 0          | 0         | 0           | 19        | 10          | 81        | 44          | <b>0.4</b> | <b>54</b>    |
| <b>Total</b>           | <b>3</b> | <b>371</b> | <b>16</b> | <b>2099</b> | <b>11</b> | <b>1474</b> | <b>70</b> | <b>9193</b> | <b>100</b> | <b>13137</b> |

Table 55: Contribution of adverbs and sentence type to the selection of *do*-support

Another factor that has been associated with the use of *do*-support is the position of the adverb relative to the verb. Ihalainen (1976: 609) links pre-verbal adverbials with the use of *do*-support in Present Day Somerset English. Pérez (1997: 41) also finds *do*-support to be favoured with pre-verbal adverbials, and suggests that this may be due to a general shift away from post-verbal adverbials in the Early Modern period. In Early Modern English data, Söderlund (2017: 269) does not find the presence of a pre-verbal adverbial to be a strong predictor for *do*-support, and Rissanen (1985) finds pre-verbal adverbials to be infrequent with *do*-support. Wischer (2008: 149), however, claims that pre-verbal adverbials do favour the use of *do*-support in Early Modern English.

Table 56 suggests that while pre-verbal adverbials \*(DO + ADV + V) and post-verbal adverbials (DO + V + ADV) do not exert much of an effect on *do*-support, cases such as (108), where there is both a pre-verbal and a post-verbal adverbial (DO + ADV + V + ADV), appear to favour the use of *do*-support (27%). Post-verbal adverbials appear with *do*-support slightly less frequently than pre-verbal, in line with claims made by Pérez (1997: 41).

(108)        and **Immediatly** gooddy Estick did sett upon me **most dreadfully**

(SWP045.16 – Elizabeth Hubbard,  
Deposition: 4<sup>th</sup> August, 1692)

| Factors      | <i>do</i> -support |     | simple |      | <b>Total</b> |             |
|--------------|--------------------|-----|--------|------|--------------|-------------|
|              | %                  | N   | %      | N    | %            | N           |
| DO + ADV + V | 14                 | 267 | 86     | 1607 | <b>77</b>    | <b>1874</b> |

|                       |    |     |    |      |            |                          |
|-----------------------|----|-----|----|------|------------|--------------------------|
| DO + V +<br>ADV       | 12 | 47  | 88 | 337  | <b>16</b>  | <b>384</b>               |
| DO + ADV +<br>V + ADV | 27 | 48  | 73 | 133  | <b>7</b>   | <b>181</b>               |
| <b>Total</b>          | 15 | 362 | 85 | 2077 | <b>100</b> | <b>2439<sup>20</sup></b> |

Table 56: Contribution of adverb position to the selection of *do*-support

In order to test the effect of an adverb intervening between *do*-support and the main verb, a subset of the data (+DO, +ADV) is represented in Table 57. As can be seen, when *do*-support and adverbs co-occur, the adverb intervenes between DO and the main verb 45% of the time. This dovetails with findings reported by Ellegård (1953: 182), who reports that adverbs in an intervening context were frequent with *do*-support (and become increasingly so throughout the period that Ellegård's data covers).

| Factors                                 | Total      |            |
|-----------------------------------------|------------|------------|
|                                         | %          | N          |
| Intervening (DO+ADV+V)                  | 45         | 163        |
| Pre-verbal (ADV+DO+V)                   | 24         | 87         |
| Multiple Adverbs (Pre- and Post-verbal) | 18         | 65         |
| Post-verbal (DO+V+ADV)                  | 13         | 47         |
| <b>Total</b>                            | <b>100</b> | <b>362</b> |

Table 57: Distribution of *do*-support by adverb placement

Table 58 cross-tabulates the effects of adverb placement and sentence type on the selection of *do*-support. Contexts with both a pre-verbal and post-verbal adverb particularly favour *do*-support in affirmative declaratives (28%). Affirmative interrogatives with an adverb favour the use of *do*-support, regardless of the placement of the adverb (49%. for pre-verbal, 52% for post-verbal), although pre- and post-adverbial

<sup>20</sup> Tokens without adverbs were omitted from this table.

contexts yield minimal tokens. Imperatives with adverbs are rare, but in affirmative contexts predominantly have a pre-verbal adverb. Such contexts categorically appear with the simple form.

Wischer (2008: 150) suggests that the use of *do*-support in negative contexts may have originated with pre-verbal adverbs. Engblom (1938) also connects the rise of negative *do*-support with the presence of adverbs. Ellegård (1953: 198) finds *do*-support in negative contexts to be favoured with pre-verbal adverbs, with post-verbal adverbials occurring quasi-categorically with the simple form. Negative interrogatives with adverbs are relatively rare in the *Salem Witchcraft Papers*, but do seem to favour the use of *do*-support, particularly with pre-verbal adverbs.<sup>21</sup> Post-verbal adverbs are infrequent in negative declaratives, and pre-verbal adverbs do not appear to favour the use of *do*-support in this context with a rate of only 16%. Recall that in Table 45, negative declaratives are shown to select *do*-support 37% of the time. If *do*-support in negatives originated with pre-verbal adverbs, it might be expected that *do*-support would be favoured in this context. It is not clear that this is the case, at least in negative declaratives.

| Factors                      | DO + ADV + V           |     |        |      | DO + V + ADV           |    |        |     | DO + ADV + V +<br>ADV  |    |        |     | Total |      |
|------------------------------|------------------------|-----|--------|------|------------------------|----|--------|-----|------------------------|----|--------|-----|-------|------|
|                              | <i>do</i> -<br>support |     | simple |      | <i>do</i> -<br>support |    | simple |     | <i>do</i> -<br>support |    | simple |     |       |      |
|                              | %                      | N   | %      | N    | %                      | N  | %      | N   | %                      | N  | %      | N   | %     | N    |
| Affirmative<br>Declarative   | 12                     | 191 | 88     | 1409 | 6                      | 18 | 94     | 309 | 28                     | 46 | 72     | 120 | 86    | 2093 |
| Affirmative<br>Interrogative | 49                     | 35  | 51     | 36   | 52                     | 22 | 48     | 20  | 33                     | 1  | 67     | 2   | 5     | 116  |
| Affirmative<br>Imperative    | 0                      | 0   | 100    | 26   | 17                     | 1  | 83     | 5   | 0                      | 0  | 100    | 1   | 1     | 33   |
|                              |                        |     |        |      |                        |    |        |     |                        |    |        |     |       |      |
| Negative<br>Declarative      | 16                     | 25  | 84     | 135  | 60                     | 3  | 40     | 2   | 10                     | 1  | 90     | 9   | 7     | 175  |

<sup>21</sup> Post-verbal adverbs in negatives are not frequent enough to be commented upon.

|                        |           |            |           |             |          |           |           |            |          |           |          |            |            |                          |
|------------------------|-----------|------------|-----------|-------------|----------|-----------|-----------|------------|----------|-----------|----------|------------|------------|--------------------------|
| Negative Interrogative | 94        | 16         | 6         | 1           | 75       | 3         | 25        | 1          | 0        | 0         | 100      | 1          | 1          | 22                       |
| <b>Total</b>           | <b>11</b> | <b>267</b> | <b>66</b> | <b>1607</b> | <b>2</b> | <b>47</b> | <b>14</b> | <b>337</b> | <b>2</b> | <b>48</b> | <b>5</b> | <b>133</b> | <b>100</b> | <b>2439<sup>22</sup></b> |

Table 58: Contribution of adverb placement and sentence type to the selection of *do*-support

#### 4.4.3 Lexical Effects

One of the more widely cited factors influencing the use of *do*-support is its use in specific lexical collocations. Of particular interest is Ellegård's (1953) *know*-group (Section 4.4.3.1).

In the *Salem Witchcraft Papers*, there are 556 different main verbs that appear (or could theoretically appear) with *do*-support. Analysis will be restricted only to the highest frequency verbs (a group of 29 verbs that each occur more than 100 times, resulting in 8769 tokens), as many of the verbs used with *do*-support occur only once or twice, and are therefore not amenable to detailed quantitative analysis. This partitioning of the data results in a grouping of 29 high-frequency verbs, given in Table 59 below. Table 59 draws on three complementary measures: the proportion each lexical item represents out of the total pool of lexical items (% pool); the strength of each lexical item's association with *do*-support, as measured by overall rate of *do*-support with the lexical item in question (rate *do*-support); and the proportion of *do*-support each lexical item represents out of all *do*-support tokens (% *do*-support).

| Lexical Item | % Pool         | Rate <i>do</i> -support | % <i>do</i> -support |
|--------------|----------------|-------------------------|----------------------|
| Torment      | 1% (107/13137) | 41% (44/107)            | 2% (44/1845)         |
| Present      | 1% (122/13137) | 39% (47/122)            | 3% (47/1845)         |
| Know         | 4% (483/13137) | 37% (178/483)           | 10% (178/1845)       |
| Hurt         | 2% (211/13137) | 34% (71/211)            | 4% (71/1845)         |
| Think        | 2% (207/13137) | 27% (56/207)            | 3% (56/1845)         |
| Look         | 1% (129/13137) | 27% (35/129)            | 2% (35/1845)         |
| Do           | 1% (145/13137) | 23% (34/145)            | 2% (34/1845)         |
| Afflict      | 2% (314/13137) | 22% (70/314)            | 4% (70/1845)         |

<sup>22</sup> Tokens without adverbs were omitted from this table.

|         |                  |               |               |
|---------|------------------|---------------|---------------|
| Believe | 1% (147/13137)   | 20% (30/147)  | 2% (30/1845)  |
| See     | 5% (698/13137)   | 18% (124/698) | 7% (124/1845) |
| Make    | 2% (234/13137)   | 16% (37/234)  | 2% (37/1845)  |
| Hear    | 1% (153/13137)   | 16% (25/153)  | 1% (25/1845)  |
| Appear  | 2% (206/13137)   | 15% (31/206)  | 2% (31/1845)  |
| Give    | 1% (126/13137)   | 15% (19/126)  | 1% (19/1845)  |
| Go      | 3% (338/13137)   | 12% (42/338)  | 2% (42/1845)  |
| Own     | 1% (175/13137)   | 12% (21/175)  | 1% (21/1845)  |
| Testify | 3% (392/13137)   | 10% (41/392)  | 2% (41/1845)  |
| Bring   | 1% (182/13137)   | 8% (16/182)   | 1% (16/1845)  |
| Ask     | 1% (162/13137)   | 8% (13/162)   | 1% (13/1845)  |
| Affirm  | 1% (144/13137)   | 8% (12/144)   | 1% (12/1845)  |
| Take    | 1% (116/13137)   | 8% (9/116)    | 0.4% (9/1845) |
| Say     | 13% (1748/13137) | 7% (119/1748) | 6% (119/1845) |
| Stand   | 1% (105/13137)   | 6% (6/105)    | 0.3% (6/1845) |
| Answer  | 1% (130/13137)   | 5% (7/130)    | 0.4% (7/1845) |
| Cry     | 1% (111/13137)   | 5% (5/111)    | 0.3% (5/1845) |
| Tell    | 6% (761/13137)   | 4% (34/761)   | 2% (34/1845)  |
| Come    | 4% (530/13137)   | 4% (22/530)   | 1% (22/1845)  |
| Fall    | 1% (166/13137)   | 2% (4/166)    | 0.2% (4/1845) |
| Have    | 3% (427/13137)   | 0% (0/427)    | 0% (0/1845)   |

Table 59: Distribution of *do*-support with high-frequency verbs (100+ tokens)

There are several interesting observations which can here be made. Recall that the overall rate for *do*-support in the *Salem Witchcraft Papers* is 14%. Half of the high-frequency verbs shown in Table 59 above disfavour the use of *do*-support. *Have* categorically, and *fall* near-categorically, appear with the simple form. Verbs indicating speech-acts (*own*, *testify*, *ask*, *affirm*, *say*, *answer*, *cry*, and *tell*) also disfavour the use of *do*-support.

*Do*-support is favoured with verbs expressing physical pain (*afflict, hurt, torment*) and mental processes (*know, think, look, believe, present*). This may be attributable to *do*-support's association with emotive contexts, and may be a harbinger of the later emphatic reading that *do*-support would acquire. Nurmi (2018) had previously observed that *do*-support was more common with verbs expressing mental processes. Warner (2017) also found that the verb *think* favours the use of *do*-support.

#### 4.4.3.1 The Know-Verbs

Ellegård (1953: 199-201), following previous work by Dietze (1895) and Engblom (1938), observes that certain verbs were relatively slow to take *do*-support in negative declaratives. This includes verbs like *know, come, speak, and go*, which are collectively referred to as the *know*-group. Contrastively, verbs including *understand, mean, say, like, and think* were found to favour the use of *do*-support in negative declaratives in Ellegård's (1953) data. Ellegård (1953: 206-207) also looks at the rates for *do*-support with certain verbs in interrogatives, and finds that *say, mean, come, go, see, like, do, and happen* infrequently appear with *do*-support. Interestingly, *know* appears to favour the use of *do*-support in this context.

In the *Salem Witchcraft Papers*, *know* is notable not only for favouring the use of *do*-support, but also in accounting for 10% of all instances of *do*-support. This is at odds with claims that *know* was slow to adopt the use of *do*-support (Ellegård, 1953; Jones & Tagliamonte, 2004; Pérez, 1997). Examination of the use of *do*-support with *know* in different sentence types reveals that *know* disfavours *do*-support in affirmative declaratives, occurring at a rate of only (4%). *Do*-support is favoured with *know* in affirmative interrogatives (60%) and negative interrogatives (88%), and slightly disfavoured in negative declaratives (46%). These findings are in line with claims made by Ellegård (1953: 199-201) that *know* favours *do*-support in interrogatives, and that negative declarative *know* is slower to adopt *do*-support.

| Factors                 | <i>do</i> -support |   | simple |     | Total |     |
|-------------------------|--------------------|---|--------|-----|-------|-----|
|                         | %                  | N | %      | N   | %     | N   |
| Affirmative Declarative | 4                  | 6 | 96     | 135 | 29    | 141 |



|                              |           |            |           |            |            |            |
|------------------------------|-----------|------------|-----------|------------|------------|------------|
| Affirmative<br>Interrogative | 60        | 33         | 40        | 22         | 11         | 55         |
|                              |           |            |           |            |            |            |
| Negative<br>Declarative      | 46        | 124        | 54        | 146        | 56         | 270        |
| Negative<br>Interrogative    | 88        | 15         | 12        | 2          | 4          | 17         |
| <b>Total</b>                 | <b>37</b> | <b>178</b> | <b>63</b> | <b>305</b> | <b>100</b> | <b>483</b> |

Table 60: Contribution of sentence type to the selection of *do*-support with the verb *know*

Table 61 below shows the distribution of *do*-support with the main verb *come* across sentence types. *Come* has been claimed to resist *do*-support in interrogatives and negative declaratives. While negative declaratives and negative interrogatives are too infrequent to admit detailed quantitative analysis, the rate of *do*-support in affirmative interrogatives with *come* (16% compared to the overall rate of 39% *do*-support in affirmative interrogatives) appears to confirm Ellegård's (1953) claims that *come* is slow to adopt *do*-support in this context. Also of note here is the strong disfavouring of *do*-support in affirmative declaratives with *come*. This is not an effect that was previously reported by Ellegård (1953).

| Factors                      | <i>do</i> -support |   | simple |     | <b>Total</b> |     |
|------------------------------|--------------------|---|--------|-----|--------------|-----|
|                              | %                  | N | %      | N   | %            | N   |
| Affirmative<br>Declarative   | 2                  | 8 | 98     | 452 | 87           | 460 |
| Affirmative<br>Interrogative | 16                 | 9 | 84     | 48  | 11           | 57  |
| Affirmative<br>Imperative    | 0                  | 0 | 100    | 4   | 1            | 4   |
|                              |                    |   |        |     |              |     |
| Negative<br>Declarative      | 50                 | 3 | 50     | 3   | 1            | 6   |

|                        |          |           |           |            |            |            |
|------------------------|----------|-----------|-----------|------------|------------|------------|
| Negative Interrogative | 67       | 2         | 33        | 1          | <b>1</b>   | <b>3</b>   |
| <b>Total</b>           | <b>4</b> | <b>22</b> | <b>96</b> | <b>508</b> | <b>100</b> | <b>530</b> |

Table 61: Contribution of sentence type to the selection of *do*-support with the verb *come*

Ellegård (1953) found main verb *do* to resist the use of *do*-support in interrogatives with a rate of 13% (17/128). Combining rates for affirmative and negative interrogatives (allowing for comparison with Ellegård, who did not distinguish between affirmative and negative interrogatives), gives a rate of 58% (26/45) in the *Salem Witchcraft Papers*. This is clearly at odds with Ellegård's (1953: 206) claims. Additionally, main verb *do* appears to strongly resist *do*-support in affirmative declaratives, near-categorically appearing with the simple form.

| Factors                   | <i>do</i> -support |           | simple    |            | <b>Total</b> |            |
|---------------------------|--------------------|-----------|-----------|------------|--------------|------------|
|                           | %                  | N         | %         | N          | %            | N          |
| Affirmative Declarative   | 1                  | 1         | 99        | 79         | <b>55</b>    | <b>80</b>  |
| Affirmative Interrogative | 56                 | 23        | 44        | 18         | <b>28</b>    | <b>41</b>  |
| Affirmative Imperative    | 0                  | 0         | 100       | 1          | <b>1</b>     | <b>1</b>   |
|                           |                    |           |           |            |              |            |
| Negative Declarative      | 37                 | 7         | 63        | 12         | <b>13</b>    | <b>19</b>  |
| Negative Interrogative    | 75                 | 3         | 25        | 1          | <b>3</b>     | <b>4</b>   |
| <b>Total</b>              | <b>23</b>          | <b>34</b> | <b>77</b> | <b>111</b> | <b>100</b>   | <b>145</b> |

Table 62: Contribution of sentence type to the selection of *do*-support with main verb *do*

The verb *see* has been claimed to resist *do*-support in interrogatives (Ellegård, 1953: 206). This is not the case in the *Salem Witchcraft Papers*, as *do*-support shows high rates in both affirmative and negative interrogatives (83% and 97% respectively). Of particular note with the verb *see* is the lower rate of negative declarative *do* (16% compared to the overall rate of 37% negative declarative *do*).

| Factors                   | <i>do</i> -support |            | simple    |            | Total      |            |
|---------------------------|--------------------|------------|-----------|------------|------------|------------|
|                           | %                  | N          | %         | N          | %          | N          |
| Affirmative Declarative   | 9                  | 48         | 91        | 491        | <b>77</b>  | <b>539</b> |
| Affirmative Interrogative | 83*                | 33         | 18*       | 7          | <b>6</b>   | <b>40</b>  |
| Affirmative Imperative    | 0                  | 0          | 100       | 3          | <b>0.4</b> | <b>3</b>   |
|                           |                    |            |           |            |            |            |
| Negative Declarative      | 16                 | 14         | 84        | 72         | <b>12</b>  | <b>86</b>  |
| Negative Interrogative    | 97                 | 29         | 3         | 1          | <b>4</b>   | <b>30</b>  |
| <b>Total</b>              | <b>18</b>          | <b>124</b> | <b>82</b> | <b>574</b> | <b>100</b> | <b>698</b> |

Table 63: Contribution of sentence type to the selection of *do*-support with the verb *see*

Another of Ellegård's (1953: 201, 206) claims is that the verb *go* is resistant to the use of *do*-support in both interrogatives and negative declaratives. Again, low token rates for negative interrogatives necessitate combining them with affirmative interrogatives, giving a rate of 92% *do*-support (N=34). These results contrast with Ellegård's (1953) claims. *Do*-support is disfavoured in negative declaratives with *go*, with a rate of (17%), in line with Ellegård (1953), but with only 6 negative declarative tokens this is not a robust finding. Affirmative declarative *do* is also disfavoured in this context (2%).

| Factors | <i>do</i> -support | simple | Total |
|---------|--------------------|--------|-------|
|---------|--------------------|--------|-------|

|                           | %         | N         | %         | N          | %          | N          |
|---------------------------|-----------|-----------|-----------|------------|------------|------------|
| Affirmative Declarative   | 2         | 7         | 98        | 287        | 87         | 294        |
| Affirmative Interrogative | 91        | 29        | 9         | 3          | 9          | 32         |
| Affirmative Imperative    | 0         | 0         | 100       | 1          | 0.3        | 1          |
|                           |           |           |           |            |            |            |
| Negative Declarative      | 17        | 1         | 83        | 5          | 2          | 6          |
| Negative Interrogative    | 100       | 5         | 0         | 0          | 1          | 5          |
| <b>Total</b>              | <b>12</b> | <b>42</b> | <b>88</b> | <b>296</b> | <b>100</b> | <b>338</b> |

Table 64: Contribution of sentence type to the selection of *do*-support with the verb *go*

The main verb *say* has been claimed to resist *do*-support in interrogatives, but favours it in negative declaratives (Ellegård, 1953: 201, 206). Low token counts hamper reliable quantitative analysis for negative declaratives, but *do*-support appears to be actually disfavoured in this context (10%). Negative interrogatives are infrequent, but it can be seen that affirmative interrogatives only slightly resist *do*-support with *say* (33% compared to the overall rate of 39% affirmative interrogatives). Also of note here is the near-categorical disfavouring of *do*-support in affirmative declaratives with *say*, accounting for only 1% of tokens.

| Factors                   | <i>do</i> -support |    | simple |      | <b>Total</b> |      |
|---------------------------|--------------------|----|--------|------|--------------|------|
|                           | %                  | N  | %      | N    | %            | N    |
| Affirmative Declarative   | 1                  | 17 | 99     | 1413 | 82           | 1430 |
| Affirmative Interrogative | 33                 | 86 | 67     | 177  | 15           | 263  |

|                           |          |            |           |             |            |             |
|---------------------------|----------|------------|-----------|-------------|------------|-------------|
| Affirmative<br>Imperative | 0        | 0          | 100       | 4           | <b>0.2</b> | <b>4</b>    |
|                           |          |            |           |             |            |             |
| Negative<br>Declarative   | 10       | 4          | 90        | 35          | <b>2</b>   | <b>39</b>   |
| Negative<br>Interrogative | 100      | 11         | 0         | 0           | <b>1</b>   | <b>11</b>   |
| Negative<br>Imperative    | 0        | 0          | 100       | 1           | <b>0</b>   | <b>1</b>    |
| <b>Total</b>              | <b>7</b> | <b>118</b> | <b>93</b> | <b>1630</b> | <b>100</b> | <b>1748</b> |

Table 65: Contribution of sentence type to the selection of *do*-support with the verb *say*

Several observations may be made about the *know*-group of verbs, as cross-tabulating these verbs with sentence type has proved illuminating. Of the 6 verbs discussed here (*know*, *come*, *do*, *see*, *go*, and *say*), only *know* and *come* align with the claims advanced by Ellegård (1953: 201, 206). One possible explanation for this divergence is that Ellegård's (1953) claims are based on a 150-year period from 1550-1700. It could be that resistance to *do*-support among these verbs was strongest in earlier texts, when change was in its earlier phases, and had ceased to be operative by the end of the 17<sup>th</sup> century, when change was more advanced.

Another key finding relating to these verbs is that most of them (with the exception of *see*) disfavour *do*-support in affirmative declaratives. This could indicate that these verbs are leading in the decline of affirmative declarative *do*, though it would first need to be shown that they were even productive with this context.

Discussion will now turn briefly to the other high-frequency verbs in the *Salem Witchcraft Papers* which are not traditionally considered among the *know*-verbs, but are nonetheless high-frequency (more than 100 instances) enough to warrant more detailed analysis. For ease of navigation, they have been grouped by shared semantic characteristics.

#### 4.4.3.2 Verbs Expressing Mental Processes

Nevalainen and Rissanen (1986: 43-44) connect the use of *do*-support in affirmative declaratives with verbs expressing mental processes. As will be shown, the data from the *Salem Witchcraft Papers* is somewhat inconsistent on this point.

The verb *think* has previously been claimed to favour *do*-support in interrogatives (Ellegård, 1953: 206), which does appear to be the case, with a rate of 76% (37/49) when affirmatives and negatives are combined.

| Factors                   | <i>do</i> -support |           | simple    |            | <b>Total</b> |            |
|---------------------------|--------------------|-----------|-----------|------------|--------------|------------|
|                           | %                  | N         | %         | N          | %            | N          |
| Affirmative Declarative   | 6                  | 9         | 94        | 137        | <b>71</b>    | <b>146</b> |
| Affirmative Interrogative | 74                 | 34        | 26        | 12         | <b>22</b>    | <b>46</b>  |
|                           |                    |           |           |            |              |            |
| Negative Declarative      | 83                 | 10        | 17        | 2          | <b>6</b>     | <b>12</b>  |
| Negative Interrogative    | 100                | 3         | 0         | 0          | <b>1</b>     | <b>3</b>   |
| <b>Total</b>              | <b>27</b>          | <b>56</b> | <b>73</b> | <b>151</b> | <b>100</b>   | <b>207</b> |

Table 66: Contribution of sentence type to the selection of *do*-support with the verb *think*

*Hear* shows little effect on the use of *do*-support in affirmative declaratives. *Hear* is infrequent outside of this context, but there are indications that it disfavours *do*-support in negative declaratives (20% compared to overall rate of 37% *do*-support in negative declaratives in Table 45), and favours *do*-support in interrogatives.

| Factors | <i>do</i> -support | simple | <b>Total</b> |
|---------|--------------------|--------|--------------|
|---------|--------------------|--------|--------------|

|                           | %         | N         | %         | N          | %          | N          |
|---------------------------|-----------|-----------|-----------|------------|------------|------------|
| Affirmative Declarative   | 8         | 10        | 92        | 111        | <b>79</b>  | <b>121</b> |
| Affirmative Interrogative | 78        | 7         | 22        | 2          | <b>6</b>   | <b>9</b>   |
|                           |           |           |           |            |            |            |
| Negative Declarative      | 20        | 4         | 80        | 16         | <b>13</b>  | <b>20</b>  |
| Negative Interrogative    | 100       | 3         | 0         | 0          | <b>2</b>   | <b>3</b>   |
| <b>Total</b>              | <b>16</b> | <b>24</b> | <b>84</b> | <b>129</b> | <b>100</b> | <b>153</b> |

Table 67: Contribution of sentence type to the selection of *do*-support with the verb *hear*

*Do*-support is favoured in affirmative declaratives (15% compared with the overall rate of 8% affirmative declarative *do*) with the verb *believe*. Token numbers outside of affirmative declaratives are too low to reach any definitive conclusions, but what little variation there is suggests a general preference for *do*-support regardless of context.

| Factors                   | <i>do</i> -support |    | simple |     | <b>Total</b> |            |
|---------------------------|--------------------|----|--------|-----|--------------|------------|
|                           | %                  | N  | %      | N   | %            | N          |
| Affirmative Declarative   | 15                 | 21 | 85     | 117 | <b>94</b>    | <b>138</b> |
| Affirmative Interrogative | 100                | 5  | 0      | 0   | <b>3</b>     | <b>5</b>   |
|                           |                    |    |        |     |              |            |
| Negative Declarative      | 100                | 3  | 0      | 0   | <b>2</b>     | <b>3</b>   |
| Negative Interrogative    | 100                | 1  | 0      | 0   | <b>1</b>     | <b>1</b>   |

|              |           |           |           |            |            |            |
|--------------|-----------|-----------|-----------|------------|------------|------------|
| <b>Total</b> | <b>20</b> | <b>30</b> | <b>80</b> | <b>117</b> | <b>100</b> | <b>147</b> |
|--------------|-----------|-----------|-----------|------------|------------|------------|

Table 68: Contribution of sentence type to the selection of *do*-support with the verb *believe*

*Look* favours the use of *do*-support in affirmative interrogatives (84% compared to 39% for the overall rate of *do*-support in this context).

| Factors                   | <i>do</i> -support |           | simple    |           | <b>Total</b> |            |
|---------------------------|--------------------|-----------|-----------|-----------|--------------|------------|
|                           | %                  | N         | %         | N         | %            | N          |
| Affirmative Declarative   | 11                 | 9         | 89        | 76        | <b>66</b>    | <b>85</b>  |
| Affirmative Interrogative | 84                 | 26        | 16        | 5         | <b>24</b>    | <b>31</b>  |
| Affirmative Imperative    | 0                  | 0         | 100       | 11        | <b>9</b>     | <b>11</b>  |
|                           |                    |           |           |           |              |            |
| Negative Declarative      | 0                  | 0         | 100       | 2         | <b>2</b>     | <b>2</b>   |
| <b>Total</b>              | <b>27</b>          | <b>35</b> | <b>73</b> | <b>94</b> | <b>100</b>   | <b>129</b> |

Table 69: Contribution of sentence type to the selection of *do*-support with the verb *look*

Despite claims in the literature that verbs expressing mental processes favour *do*-support in affirmative declaratives in Modern English (Nevalainen & Rissanen, 1986), *believe* is the only high-frequency verb to actually favour affirmative declarative *do*, and even with *believe* it is only a rate of 15%. The other verbs expressing mental processes show little to no effect on the use of *do*-support in affirmative declaratives.

Another observable pattern here is that these verbs tend to favour *do*-support in interrogatives. Verbs like *think*, *look*, and *hear* all appear to favour *do*-support in interrogatives. *Know* and *see* would also fit with this group, both as cognitive-state verbs, and in terms of rates of *do*-support in interrogatives.



#### 4.4.3.3 Verbs Expressing Speech Acts

A pattern that is observable in the data here is that verbs expressing speech acts generally disfavour *do*-support in affirmative declaratives.

*Tell* is generally rare in negative contexts (but does seem to favour *do*-support here), so discussion will focus on the affirmative, where *do*-support is disfavoured. *Do*-support is disfavoured in affirmative declaratives (1%) and affirmative imperatives (2%).

Most notable is the rate of 19% *do*-support in affirmative interrogatives, a marked departure from the overall rate of 39% *do*-support in this context (Table 45).

| Factors                   | <i>do</i> -support |           | simple    |            | Total      |            |
|---------------------------|--------------------|-----------|-----------|------------|------------|------------|
|                           | %                  | N         | %         | N          | %          | N          |
| Affirmative Declarative   | 1                  | 8         | 99        | 628        | 84         | 636        |
| Affirmative Interrogative | 19                 | 12        | 81        | 50         | 8          | 62         |
| Affirmative Imperative    | 2                  | 1         | 98        | 46         | 6          | 47         |
|                           |                    |           |           |            |            |            |
| Negative Declarative      | 71                 | 5         | 29        | 2          | 1          | 7          |
| Negative Interrogative    | 100                | 7         | 0         | 0          | 1          | 7          |
| Negative Imperative       | 50                 | 1         | 50        | 1          | 0.2        | 2          |
| <b>Total</b>              | <b>4</b>           | <b>34</b> | <b>96</b> | <b>727</b> | <b>100</b> | <b>761</b> |

Table 70: Contribution of sentence type to the selection of *do*-support with the verb *tell*

The verb *ask* is predominantly found in affirmative declaratives, where it disfavours the use of *do*-support (3%). Other contexts are too infrequent to warrant consideration.

| Factors                   | <i>do</i> -support |           | Simple    |            | Total      |            |
|---------------------------|--------------------|-----------|-----------|------------|------------|------------|
|                           | %                  | N         | %         | N          | %          | N          |
| Affirmative Declarative   | 3                  | 5         | 97        | 144        | 92         | 149        |
| Affirmative Interrogative | 43                 | 3         | 57        | 4          | 4          | 7          |
|                           |                    |           |           |            |            |            |
| Negative Declarative      | 100                | 3         | 0         | 0          | 2          | 3          |
| Negative Interrogative    | 50                 | 1         | 50        | 1          | 1          | 2          |
| Negative Imperative       | 100                | 1         | 0         | 0          | 1          | 1          |
| <b>Total</b>              | <b>8</b>           | <b>13</b> | <b>92</b> | <b>149</b> | <b>100</b> | <b>162</b> |

Table 71: Contribution of sentence type to the selection of *do*-support with the verb *ask*

The verb *answer* shows a strong aversion to *do*-support in affirmative declaratives (1%), near-categorically appearing with the simple form.

| Factors                   | <i>do</i> -support |   | Simple |     | Total |     |
|---------------------------|--------------------|---|--------|-----|-------|-----|
|                           | %                  | N | %      | N   | %     | N   |
| Affirmative Declarative   | 1                  | 1 | 99     | 119 | 92    | 120 |
| Affirmative Interrogative | 50                 | 1 | 50     | 1   | 2     | 2   |

|                        |          |          |           |            |            |            |
|------------------------|----------|----------|-----------|------------|------------|------------|
| Affirmative Imperative | 0        | 0        | 100       | 1          | <b>1</b>   | <b>1</b>   |
|                        |          |          |           |            |            |            |
| Negative Declarative   | 60       | 3        | 40        | 2          | <b>4</b>   | <b>5</b>   |
| Negative Interrogative | 100      | 2        | 0         | 0          | <b>2</b>   | <b>2</b>   |
| <b>Total</b>           | <b>5</b> | <b>7</b> | <b>95</b> | <b>123</b> | <b>100</b> | <b>130</b> |

Table 72: Contribution of sentence type to the selection of *do*-support with the verb *answer*

The verb *cry* disfavours *do*-support in affirmative declaratives (4%), and is almost entirely absent from other contexts.

| Factors                   | <i>do</i> -support |          | Simple    |            | Total      |            |
|---------------------------|--------------------|----------|-----------|------------|------------|------------|
|                           | %                  | N        | %         | N          | %          | N          |
| Affirmative Declarative   | 4                  | 4        | 96        | 101        | <b>95</b>  | <b>105</b> |
| Affirmative Interrogative | 0                  | 0        | 100       | 5          | <b>5</b>   | <b>5</b>   |
|                           |                    |          |           |            |            |            |
| Negative Declarative      | 0                  | 0        | 100       | 1          | <b>1</b>   | <b>1</b>   |
| <b>Total</b>              | <b>4</b>           | <b>4</b> | <b>96</b> | <b>107</b> | <b>100</b> | <b>111</b> |

Table 73: Contribution of sentence type to the selection of *do*-support with the verb *cry*

Verbs expressing speech acts (*tell*, *ask*, *answer*, and *cry*) are consistent in disfavoursing *do*-support in affirmative declaratives. Of the *know*-verbs, *say* would belong to this group, and patterns accordingly, disfavoursing affirmative declarative *do*.

#### 4.4.3.4 Emotive Verbs

The emotive nature of *do*-support has been frequently claimed in the literature (Bækken, 2002; Ellegård, 1953), and so it might be expected that verbs imbued with emotive connotations might favour the use of *do*-support. Among the high-frequency verbs, *afflict*, *hurt*, and *torment* stand out as particularly emotive considering the socio-historical context, as they were often used when the accusers were discussing how the accused affected them with witchcraft (Examples 109 and 110 below).

(109) S<u>zana shelten saith that wife buckle **did** Aflicke me

(SWP065.2 – Susannah Sheldon,  
Complaint: 17<sup>th</sup> May, 1692)

(110) in the time of hir Examination she **did** torment me most dreadfully by biting pinching  
and almost choaking me

(SWP038.14 – Elizabeth Hubbard,  
Deposition: 4<sup>th</sup> August, 1692)

*Do*-support is favoured in affirmative declaratives (17%) with *afflict*, the only context where this verb is particularly frequent.

| Factors                   | <i>do</i> -support |    | simple |     | Total     |            |
|---------------------------|--------------------|----|--------|-----|-----------|------------|
|                           | %                  | N  | %      | N   | %         | N          |
| Affirmative Declarative   | 17                 | 47 | 83     | 227 | <b>87</b> | <b>274</b> |
| Affirmative Interrogative | 61                 | 17 | 39     | 11  | <b>9</b>  | <b>28</b>  |
|                           |                    |    |        |     |           |            |
| Negative Declarative      | 33                 | 3  | 67     | 6   | <b>3</b>  | <b>9</b>   |

|                        |           |           |           |            |            |            |
|------------------------|-----------|-----------|-----------|------------|------------|------------|
| Negative Interrogative | 100       | 3         | 0         | 0          | 1          | 3          |
| <b>Total</b>           | <b>22</b> | <b>70</b> | <b>78</b> | <b>244</b> | <b>100</b> | <b>314</b> |

Table 74: Contribution of sentence type to the selection of *do*-support with the verb *afflict*

*Hurt* shows a slightly higher rate of *do*-support in affirmative declaratives (14%) than the overall rate for this context (8%, per Table 45). This is also the case for affirmative interrogatives (48%) and negative declaratives (52%).

| Factors                   | <i>do</i> -support |           | simple    |            | Total      |            |
|---------------------------|--------------------|-----------|-----------|------------|------------|------------|
|                           | %                  | N         | %         | N          | %          | N          |
| Affirmative Declarative   | 14                 | 14        | 86        | 83         | 46         | 97         |
| Affirmative Interrogative | 48                 | 34        | 52        | 37         | 34         | 71         |
| Affirmative Imperative    | 0                  | 0         | 100       | 5          | 2          | 5          |
|                           |                    |           |           |            |            |            |
| Negative Declarative      | 52                 | 15        | 48        | 14         | 14         | 29         |
| Negative Interrogative    | 89                 | 8         | 11        | 1          | 4          | 9          |
| <b>Total</b>              | <b>34</b>          | <b>71</b> | <b>66</b> | <b>140</b> | <b>100</b> | <b>211</b> |

Table 75: Contribution of sentence type to the selection of *do*-support with the verb *hurt*

*Torment* shows a notably high rate of *do*-support in affirmative declaratives (40%).

| Factors | <i>do</i> -support |   | simple |   | Total |   |
|---------|--------------------|---|--------|---|-------|---|
|         | %                  | N | %      | N | %     | N |

|                           |           |           |           |           |            |            |
|---------------------------|-----------|-----------|-----------|-----------|------------|------------|
| Affirmative Declarative   | 40        | 39        | 60        | 59        | 92         | 98         |
| Affirmative Interrogative | 50        | 4         | 50        | 4         | 7          | 8          |
|                           |           |           |           |           |            |            |
| Negative Declarative      | 100       | 1         | 0         | 0         | 1          | 1          |
| <b>Total</b>              | <b>41</b> | <b>44</b> | <b>59</b> | <b>63</b> | <b>100</b> | <b>107</b> |

Table 76: Contribution of sentence type to the selection of *do*-support with the verb *torment*

The high-frequency emotive verbs in the data do seem to follow a general trend, whereby *do*-support is favoured in affirmative declarative contexts. *Hurt* in particular appears to generally favour *do*-support across sentence types. The concept of entrenchment, as discussed by Bybee and Beckner (2010: 840), could go some way to explaining the favouring of *do*-support with these high-frequency verbs. Entrenchment is a process whereby high-frequency constructions become ‘entrenched’ as a single unit (Bybee & Beckner, 2010: 840). These entrenched collocations are more resistant to change, which in this case could mean that high-frequency verbs in the data resist the decline of *do*-support in affirmative declaratives.

#### 4.4.3.5 Legal Terminology

Another possible pattern is the use of certain verbs belonging to legal discourse. The use of these verbs is highly context-dependent, and it might be expected that this would be reflected in the rates for *do*-support.

The verb *testify* favours the use of *do*-support in affirmative declaratives with a rate of 16%. This could reflect its use in formulaic legal language (as in ‘*I do testify that...*’). The low rate of affirmative interrogative *do* with this verb (1%) could support this assertion.

|         |                    |        |              |
|---------|--------------------|--------|--------------|
| Factors | <i>do</i> -support | simple | <b>Total</b> |
|---------|--------------------|--------|--------------|

|                           | %         | N         | %         | N          | %          | N          |
|---------------------------|-----------|-----------|-----------|------------|------------|------------|
| Affirmative Declarative   | 16        | 40        | 84        | 208        | 63         | 248        |
| Affirmative Interrogative | 1         | 1         | 99        | 143        | 37         | 144        |
| <b>Total</b>              | <b>10</b> | <b>41</b> | <b>90</b> | <b>351</b> | <b>100</b> | <b>392</b> |

Table 77: Contribution of sentence type to the selection of do-support with the verb testify

Affirmative declarative *do* is slightly favoured (12%) with the verb *own* (in the sense of ‘admit,’ ‘acknowledge’) in the context ‘*She owned her confession to be true*’.

| Factors                   | <i>do</i> -support |           | simple    |            | Total      |            |
|---------------------------|--------------------|-----------|-----------|------------|------------|------------|
|                           | %                  | N         | %         | N          | %          | N          |
| Affirmative Declarative   | 12                 | 20        | 88        | 148        | 96         | 168        |
| Affirmative Interrogative | 0                  | 0         | 100       | 1          | 1          | 1          |
| Affirmative Imperative    | 100                | 2         | 0         | 0          | 1          | 2          |
|                           |                    |           |           |            |            |            |
| Negative Declarative      | 0                  | 0         | 100       | 3          | 2          | 3          |
| Negative Interrogative    | 100                | 1         | 0         | 0          | 1          | 1          |
| <b>Total</b>              | <b>13</b>          | <b>23</b> | <b>87</b> | <b>152</b> | <b>100</b> | <b>175</b> |

Table 78: Contribution of sentence type to the selection of do-support with the verb own

The verb *present* is almost exclusively found in affirmative declaratives, where it strongly favours the use of *do*-support with a rate of 38%. This may be context-dependent, with the presentation of evidence in a legal setting.

| Factors                   | <i>do</i> -support |           | simple    |           | <b>Total</b> |            |
|---------------------------|--------------------|-----------|-----------|-----------|--------------|------------|
|                           | %                  | N         | %         | N         | %            | N          |
| Affirmative Declarative   | 38                 | 45        | 63        | 75        | <b>98</b>    | <b>120</b> |
| Affirmative Interrogative | 100                | 2         | 0         | 0         | <b>2</b>     | <b>2</b>   |
| <b>Total</b>              | <b>39</b>          | <b>47</b> | <b>61</b> | <b>75</b> | <b>100</b>   | <b>122</b> |

Table 79: Contribution of sentence type to the selection of *do*-support with the verb *present*

The verb *affirm* is almost entirely restricted to affirmative declaratives, but does not show any notable influence on the use of *do*-support, as rates are in line with the overall rate of *do*-support in that context (8% in Table 45).

| Factors                   | <i>do</i> -support |           | simple    |            | <b>Total</b> |            |
|---------------------------|--------------------|-----------|-----------|------------|--------------|------------|
|                           | %                  | N         | %         | N          | %            | N          |
| Affirmative Declarative   | 8                  | 12        | 92        | 131        | <b>99</b>    | <b>143</b> |
| Affirmative Interrogative | 0                  | 0         | 100       | 1          | <b>1</b>     | <b>1</b>   |
| <b>Total</b>              | <b>8</b>           | <b>12</b> | <b>92</b> | <b>132</b> | <b>100</b>   | <b>144</b> |

Table 80: Contribution of sentence type to the selection of *do*-support with the verb *affirm*

With the exception of *affirm*, the collection of verbs that relate most closely to the legal context of the *Salem Witch Trials* generally favour *do*-support in affirmative declaratives. This could reflect style and register effects of the legal context.



#### 4.4.3.6 Other High-Frequency Verbs

Certain high-frequency verbs found in the data are not as easily categorised, but are still worth discussing. For example, the verb *fall* near-categorically takes the simple form in affirmative declaratives (and is rare outside of this context).

| Factors                   | <i>do</i> -support |          | simple    |            | <b>Total</b> |            |
|---------------------------|--------------------|----------|-----------|------------|--------------|------------|
|                           | %                  | N        | %         | N          | %            | N          |
| Affirmative Declarative   | 2                  | 3        | 98        | 160        | <b>98</b>    | <b>163</b> |
| Affirmative Interrogative | 0                  | 0        | 100       | 1          | <b>1</b>     | <b>1</b>   |
|                           |                    |          |           |            |              |            |
| Negative Declarative      | 0                  | 0        | 100       | 1          | <b>1</b>     | <b>1</b>   |
| Negative Interrogative    | 100                | 1        | 0         | 0          | <b>1</b>     | <b>1</b>   |
| <b>Total</b>              | <b>2</b>           | <b>4</b> | <b>98</b> | <b>162</b> | <b>100</b>   | <b>166</b> |

Table 81: Contribution of sentence type to the selection of *do*-support with the verb *fall*

Similarly, at no point in the data does the verb *have* appear with *do*-support, even in negatives and interrogatives. This is a notable finding that does not appear to have been reported previously, and is clearly at odds with the status of *do*-support in Present Day English, where phrases like ‘*I don’t have it*’ are preferred to the archaic simple alternative ‘*I have it not*’.

| Factors                 | <i>do</i> -support |   | simple |     | <b>Total</b> |            |
|-------------------------|--------------------|---|--------|-----|--------------|------------|
|                         | %                  | N | %      | N   | %            | N          |
| Affirmative Declarative | 0                  | 0 | 100    | 307 | <b>71</b>    | <b>307</b> |

|                           |          |          |            |            |            |            |
|---------------------------|----------|----------|------------|------------|------------|------------|
| Affirmative Interrogative | 0        | 0        | 100        | 41         | <b>10</b>  | <b>41</b>  |
|                           |          |          |            |            |            |            |
| Negative Declarative      | 0        | 0        | 100        | 73         | <b>17</b>  | <b>73</b>  |
| Negative Interrogative    | 0        | 0        | 100        | 6          | <b>1</b>   | <b>6</b>   |
| <b>Total</b>              | <b>0</b> | <b>0</b> | <b>100</b> | <b>427</b> | <b>100</b> | <b>427</b> |

Table 82: Contribution of sentence type to the selection of *do*-support with the verb *have*

*Make* is frequent, but does not display much deviation from the general rates for *do*-support across sentence types. Affirmative declarative *do* with this verb is only slightly more frequent than the average (14% compared to 9%).

| Factors                   | <i>do</i> -support |           | simple    |            | <b>Total</b> |            |
|---------------------------|--------------------|-----------|-----------|------------|--------------|------------|
|                           | %                  | N         | %         | N          | %            | N          |
| Affirmative Declarative   | 14                 | 24        | 86        | 152        | <b>75</b>    | <b>176</b> |
| Affirmative Interrogative | 40                 | 12        | 60        | 18         | <b>13</b>    | <b>30</b>  |
| Affirmative Imperative    | 0                  | 0         | 100       | 21         | <b>9</b>     | <b>21</b>  |
|                           |                    |           |           |            |              |            |
| Negative Declarative      | 0                  | 0         | 100       | 6          | <b>3</b>     | <b>6</b>   |
| Negative Interrogative    | 100                | 1         | 0         | 0          | <b>0.4</b>   | <b>1</b>   |
| <b>Total</b>              | <b>16</b>          | <b>37</b> | <b>84</b> | <b>197</b> | <b>100</b>   | <b>234</b> |

Table 83: Contribution of sentence type to the selection of *do*-support with the verb *make*

*Appear* has no noticeable effect on the use of *do*-support in affirmative declaratives, but appears to favour *do*-support in affirmative interrogatives (with the caveat that this is based on only 23 tokens).

| Factors                   | <i>do</i> -support |           | simple    |            | Total      |            |
|---------------------------|--------------------|-----------|-----------|------------|------------|------------|
|                           | %                  | N         | %         | N          | %          | N          |
| Affirmative Declarative   | 9                  | 15        | 91        | 161        | <b>85</b>  | <b>176</b> |
| Affirmative Interrogative | 52                 | 12        | 48        | 11         | <b>11</b>  | <b>23</b>  |
|                           |                    |           |           |            |            |            |
| Negative Declarative      | 50                 | 3         | 50        | 3          | <b>3</b>   | <b>6</b>   |
| Negative Interrogative    | 100                | 1         | 0         | 0          | <b>0.4</b> | <b>1</b>   |
| <b>Total</b>              | <b>15</b>          | <b>31</b> | <b>85</b> | <b>175</b> | <b>100</b> | <b>206</b> |

Table 84: Contribution of sentence type to the selection of *do*-support with the verb *appear*

The verb *bring* resists *do*-support in affirmative declaratives with a rate of only 1%.

| Factors                   | <i>do</i> -support |    | simple |     | Total     |            |
|---------------------------|--------------------|----|--------|-----|-----------|------------|
|                           | %                  | N  | %      | N   | %         | N          |
| Affirmative Declarative   | 1                  | 2  | 99     | 148 | <b>82</b> | <b>150</b> |
| Affirmative Interrogative | 44                 | 11 | 56     | 14  | <b>14</b> | <b>25</b>  |
| Affirmative Imperative    | 0                  | 0  | 100    | 2   | <b>1</b>  | <b>2</b>   |
|                           |                    |    |        |     |           |            |

|                        |          |           |           |            |            |            |
|------------------------|----------|-----------|-----------|------------|------------|------------|
| Negative Declarative   | 33       | 1         | 67        | 2          | <b>2</b>   | <b>3</b>   |
| Negative Interrogative | 100      | 2         | 0         | 0          | <b>1</b>   | <b>2</b>   |
| <b>Total</b>           | <b>9</b> | <b>16</b> | <b>91</b> | <b>166</b> | <b>100</b> | <b>182</b> |

Table 85: Contribution of sentence type to the selection of *do*-support with the verb *bring*

The verb *give* shows no appreciable effect on rates of *do*-support in affirmative declaratives, but slightly favours *do*-support in interrogatives.

| Factors                   | <i>do</i> -support |           | simple    |            | <b>Total</b> |            |
|---------------------------|--------------------|-----------|-----------|------------|--------------|------------|
|                           | %                  | N         | %         | N          | %            | N          |
| Affirmative Declarative   | 9                  | 8         | 91        | 81         | <b>71</b>    | <b>89</b>  |
| Affirmative Interrogative | 46                 | 6         | 54        | 7          | <b>10</b>    | <b>13</b>  |
| Affirmative Imperative    | 0                  | 0         | 100       | 11         | <b>9</b>     | <b>11</b>  |
|                           |                    |           |           |            |              |            |
| Negative Declarative      | 20                 | 2         | 80        | 8          | <b>8</b>     | <b>10</b>  |
| Negative Interrogative    | 100                | 3         | 0         | 0          | <b>2</b>     | <b>3</b>   |
| <b>Total</b>              | <b>15</b>          | <b>19</b> | <b>85</b> | <b>107</b> | <b>100</b>   | <b>126</b> |

Table 86: Contribution of sentence type to the selection of *do*-support with the verb *give*

The verb *take* disfavours the use of affirmative declarative *do* (2%).

| Factors | <i>do</i> -support | simple | <b>Total</b> |
|---------|--------------------|--------|--------------|
|---------|--------------------|--------|--------------|

|                           | %        | N        | %         | N          | %          | N          |
|---------------------------|----------|----------|-----------|------------|------------|------------|
| Affirmative Declarative   | 2        | 2        | 98        | 102        | <b>90</b>  | <b>104</b> |
| Affirmative Interrogative | 80       | 4        | 20        | 1          | <b>4</b>   | <b>5</b>   |
| Affirmative Imperative    | 0        | 0        | 100       | 4          | <b>3</b>   | <b>4</b>   |
|                           |          |          |           |            |            |            |
| Negative Interrogative    | 100      | 3        | 0         | 0          | <b>3</b>   | <b>3</b>   |
| <b>Total</b>              | <b>8</b> | <b>9</b> | <b>93</b> | <b>108</b> | <b>100</b> | <b>116</b> |

Table 87: Contribution of sentence type to the selection of *do*-support with the verb *take*

Finally, the verb *stand* shows no appreciable effect on rates for affirmative declarative *do*, but does appear to disfavour *do*-support in affirmative interrogatives with a rate of 2%.

| Factors                   | <i>do</i> -support |          | simple    |           | <b>Total</b> |            |
|---------------------------|--------------------|----------|-----------|-----------|--------------|------------|
|                           | %                  | N        | %         | N         | %            | N          |
| Affirmative Declarative   | 8                  | 5        | 92        | 55        | <b>57</b>    | <b>60</b>  |
| Affirmative Interrogative | 2                  | 1        | 98        | 44        | <b>43</b>    | <b>45</b>  |
| <b>Total</b>              | <b>6</b>           | <b>6</b> | <b>94</b> | <b>99</b> | <b>100</b>   | <b>105</b> |

Table 88: Contribution of sentence type to the selection of *do*-support with the verb *stand*

#### 4.4.3.7 Germanic vs. Latinate Verbs

It may also prove valuable to examine the effect of verb language of origin on the use of *do*-support. This follows a claim made by Samuels (1972: 174) that verbs of Latinate origin favour *do*-support in affirmative declaratives. To this end, verbs were coded as being either Germanic or Latinate in origin, based on their entries on etymonline.com and the Oxford English Dictionary (OED). Verbs where the

origin is unclear (for example, *jump*) were omitted from this analysis. Table 89 below appears to confirm the claim made by Samuels (1972: 174); Latinate verbs in affirmative declaratives take *do*-support 14% of the time, compared to 5% for Germanic verbs. It is interesting to note that this appears to mark a more general preference for *do*-support with Latinate verbs in declaratives, as there is a higher rate of negative declarative *do* with Latinate verbs (46%) than Germanic verbs (34%). Additionally, Latinate verbs disfavour *do*-support in interrogatives, particularly in affirmative contexts (30% compared to 41% for Germanic verbs).

| Factors                   | Germanic           |             |           |             | Latinate           |            |           |             | Total      |              |
|---------------------------|--------------------|-------------|-----------|-------------|--------------------|------------|-----------|-------------|------------|--------------|
|                           | <i>do</i> -support |             | simple    |             | <i>do</i> -support |            | simple    |             |            |              |
|                           | %                  | N           | %         | N           | %                  | N          | %         | N           | %          | N            |
| Affirmative Declarative   | 5                  | 363         | 95        | 7373        | 14                 | 360        | 84        | 2242        | 80         | 10338        |
| Affirmative Interrogative | 41                 | 460         | 59        | 653         | 30                 | 111        | 70        | 257         | 11         | 1481         |
| Affirmative Imperative    | 1                  | 1           | 99        | 172         | 0                  | 0          | 100       | 26          | 2          | 199          |
|                           |                    |             |           |             |                    |            |           |             |            |              |
| Negative Declarative      | 34                 | 234         | 66        | 449         | 46                 | 43         | 54        | 51          | 6          | 777          |
| Negative Interrogative    | 90                 | 126         | 10        | 14          | 87                 | 27         | 13        | 4           | 1          | 171          |
| Negative Imperative       | 60                 | 6           | 40        | 4           | 100                | 3          | 0         | 0           | 0.1        | 13           |
| <b>Total</b>              | <b>9</b>           | <b>1190</b> | <b>67</b> | <b>8665</b> | <b>4</b>           | <b>544</b> | <b>20</b> | <b>2580</b> | <b>100</b> | <b>12979</b> |

Table 89: Contribution of sentence type and verb language of origin to the selection of *do*-support

#### 4.4.3.8 Summary of Lexical Effects

While the effects of the *know*-group first cited by Ellegård (1953) have been widely reported, and sometimes uncritically accepted, the results reported here (summarised in Table 90 below; pertinent cells are shaded for ease of reference) suggest that dividing high-frequency verbs along semantic lines can reveal certain patterns. Verbs expressing mental processes typically favour *do*-support in interrogatives (while showing very little effect in affirmative declaratives); verbs expressing speech acts consistently disfavour *do*-support in affirmative declaratives; emotive verbs favour *do*-support in affirmative declaratives, as do verbs related to legal terminology. Such consistency could suggest that the relationship

between *do*-support and specific verb collocations may not be as idiosyncratic as previously reported (Jones & Tagliamonte, 2004: 102).

| Factors                    |         | Affirmative Declaratives |         | Negative Declaratives |         | Interrogatives |        | Total |      |
|----------------------------|---------|--------------------------|---------|-----------------------|---------|----------------|--------|-------|------|
|                            |         | %                        | N       | %                     | N       | %              | N      | %     | N    |
| Mental Processes           | Know    | 4                        | 6/141   | 46                    | 124/270 | 67             | 48/72  | 6     | 483  |
|                            | See     | 9                        | 48/539  | 16                    | 14/86   | 89             | 62/70  | 8     | 695  |
|                            | Think   | 6                        | 9/146   | 83                    | 10/12   | 76             | 37/49  | 2     | 207  |
|                            | Hear    | 8                        | 10/121  | 20                    | 4/20    | 83             | 10/12  | 2     | 153  |
|                            | Believe | 15                       | 21/138  | 100                   | 3/3     | 100            | 6/6    | 2     | 147  |
|                            | Look    | 11                       | 9/85    | 0                     | 0/2     | 84             | 26/31  | 1     | 118  |
|                            |         |                          |         |                       |         |                |        |       |      |
| Speech Acts                | Say     | 1                        | 17/1430 | 10                    | 4/39    | 35             | 97/274 | 20    | 1743 |
|                            | Tell    | 1                        | 8/636   | 71                    | 5/7     | 28             | 19/69  | 8     | 712  |
|                            | Ask     | 3                        | 5/149   | 100                   | 3/3     | 44             | 4/9    | 2     | 161  |
|                            | Answer  | 1                        | 1/120   | 60                    | 3/5     | 75             | 3/4    | 1     | 129  |
|                            | Cry     | 4                        | 4/105   | 0                     | 0/1     | 0              | 0/5    | 1     | 111  |
|                            |         |                          |         |                       |         |                |        |       |      |
| Emotive                    | Afflict | 17                       | 47/274  | 33                    | 3/9     | 65             | 20/31  | 4     | 314  |
|                            | Hurt    | 14                       | 14/97   | 52                    | 15/29   | 53             | 42/80  | 2     | 206  |
|                            | Torment | 40                       | 39/98   | 100                   | 1/1     | 50             | 4/8    | 1     | 107  |
|                            |         |                          |         |                       |         |                |        |       |      |
| Legal Terminology          | Testify | 16                       | 40/248  | -                     | -       | 1              | 1/144  | 5     | 392  |
|                            | Own     | 12                       | 20/168  | 0                     | 0/3     | 50             | 1/2    | 2     | 173  |
|                            | Present | 38                       | 45/120  | -                     | -       | 100            | 2/2    | 1     | 122  |
|                            | Affirm  | 8                        | 12/143  | -                     | -       | 0              | 0/1    | 2     | 144  |
|                            |         |                          |         |                       |         |                |        |       |      |
| Other High Frequency Verbs | Come    | 2                        | 8/460   | 50                    | 3/6     | 18             | 11/60  | 6     | 526  |
|                            | Do      | 1                        | 1/80    | 37                    | 7/19    | 58             | 26/45  | 2     | 144  |
|                            | Go      | 2                        | 7/294   | 17                    | 1/6     | 92             | 34/37  | 4     | 337  |
|                            | Fall    | 2                        | 3/163   | 0                     | 0/1     | 50             | 1/2    | 2     | 166  |
|                            | Have    | 0                        | 0/307   | 0                     | 0/73    | 0              | 0/47   | 5     | 427  |
|                            | Make    | 14                       | 24/176  | 0                     | 0/6     | 42             | 13/31  | 2     | 213  |
|                            | Appear  | 9                        | 15/176  | 50                    | 3/6     | 54             | 13/24  | 2     | 206  |
|                            | Bring   | 1                        | 2/150   | 33                    | 1/3     | 48             | 13/27  | 2     | 180  |
|                            | Give    | 9                        | 8/89    | 20                    | 2/10    | 56             | 9/16   | 1     | 115  |
|                            | Take    | 2                        | 2/104   | -                     | -       | 88             | 7/8    | 1     | 112  |

|              |       |           |             |          |            |           |             |            |             |
|--------------|-------|-----------|-------------|----------|------------|-----------|-------------|------------|-------------|
|              | Stand | 8         | 5/60        | -        | -          | 2         | 1/45        | 1          | 105         |
| <b>Total</b> |       | <b>79</b> | <b>6817</b> | <b>7</b> | <b>620</b> | <b>14</b> | <b>1211</b> | <b>100</b> | <b>8648</b> |

Table 90: Summary of lexical effects on the selection of *do*-support

Grouping verbs by their language of origin (Germanic or Latinate) also proves illuminating, with Latinate verbs favouring *do*-support in declaratives and disfavouring it in interrogatives. This is congruent with the findings in Table 47, where verbs with weak morphology also favour the use of *do*-support in affirmative declaratives. The relationship between verb origin and *do*-support is not one that appears to have been investigated, beyond the claims made by Samuels (1972: 173-174), who anecdotally observes that *do*-support is more common with Latinate verbs as a method of resolving instabilities in Early Modern English verbal marking.

#### 4.4.4 External Factors

##### 4.4.4.1 Sex

Recall that Nurmi (2018: 129) claims that while female speakers appear to have been responsible for the rise in the use of *do*-support in the 16<sup>th</sup> and 17<sup>th</sup> centuries, by the end of the 17<sup>th</sup> century sex is reported to have ceased to influence the use of *do*-support. Table 91 suggests sex is not a determinant of rates of *do*-support. There appears to be almost no distinction in the rates of *do*-support between male and female speakers.

| Factors      | <i>do</i> -support |             | simple    |              | Total      |              |
|--------------|--------------------|-------------|-----------|--------------|------------|--------------|
|              | %                  | N           | %         | N            | %          | N            |
| Male         | 12                 | 753         | 88        | 5498         | 48         | 6251         |
| Female       | 13* <sup>23</sup>  | 610         | 88*       | 4270         | 37         | 4880         |
| Unknown      | 24                 | 482         | 76        | 1524         | 15         | 2006         |
| <b>Total</b> | <b>14</b>          | <b>1845</b> | <b>86</b> | <b>11292</b> | <b>100</b> | <b>13137</b> |

Table 91: Contribution of speaker sex to the selection of *do*-support

<sup>23</sup> An asterisk is included to indicate that the total percentage (13% + 88%) surpasses 100% due to rounding.



However, breaking the results down by sentence type demonstrates that there are in fact sex differences which would otherwise have remained hidden. Female speakers make less use of affirmative interrogative *do*, with a rate of only 19% (compared to 37% for male speakers), while male speakers use *do*-support in negative declaratives less frequently than female speakers (28% compared to 41% for female speakers). It is worth noting that this may be for pragmatic reasons, considering the roles typically played by male and female speakers in the trials, where female speakers were often accused, and the court officials (who are asking many of the questions during examinations) were exclusively male.

It is in fact the rates for affirmative declaratives (by far the most common construction in the data) that show no sex differentiation. The high token frequency of affirmative declaratives effectively swamps the effects of other sentence types on *do*-support when the data are examined in the aggregate. This can be resolved by cross-tabulating each factor group with sentence type, which suggests that sex effects have not entirely disappeared by 1692. This highlights a methodological confound in Nurmi (2018), who makes the claim that sex does not influence the use of *do*-support, but only looks at affirmative declaratives. As Table 92 shows, affirmative declaratives do not display sex differentiation in the use of *do*-support.

| Factors                   | Male               |     |        |      | Female             |     |        |      | Total     |             |
|---------------------------|--------------------|-----|--------|------|--------------------|-----|--------|------|-----------|-------------|
|                           | <i>do</i> -support |     | simple |      | <i>do</i> -support |     | simple |      |           |             |
|                           | %                  | N   | %      | N    | %                  | N   | %      | N    | %         | N           |
| Affirmative Declarative   | 7                  | 367 | 93     | 4776 | 8                  | 315 | 92     | 3584 | <b>81</b> | <b>9042</b> |
| Affirmative Interrogative | 37                 | 230 | 63     | 391  | 19                 | 84  | 81     | 367  | <b>10</b> | <b>1072</b> |
| Affirmative Imperative    | 1                  | 1   | 99     | 77   | 0                  | 0   | 100    | 52   | <b>1</b>  | <b>130</b>  |
|                           |                    |     |        |      |                    |     |        |      |           |             |
| Negative Declarative      | 28                 | 81  | 72     | 210  | 41                 | 185 | 59     | 263  | <b>7</b>  | <b>739</b>  |

|                        |           |            |           |             |           |            |           |             |            |              |
|------------------------|-----------|------------|-----------|-------------|-----------|------------|-----------|-------------|------------|--------------|
| Negative Interrogative | 92        | 70         | 8         | 6           | 92        | 22         | 8         | 2           | 1          | 100          |
| Negative Imperative    | 10        | 4          | 90        | 38          | 67        | 4          | 33        | 2           | 0.4        | 48           |
| <b>Total</b>           | <b>12</b> | <b>753</b> | <b>88</b> | <b>5498</b> | <b>13</b> | <b>610</b> | <b>88</b> | <b>4270</b> | <b>100</b> | <b>11131</b> |

Table 92: Contribution of speaker sex and sentence type to the selection of *do*-support

#### 4.4.4.2 Age

Warner (2005: 268-270) finds that in the period 1600-1710, age is a significant predictor of *do*-support use, with older speakers using *do*-support in negative declaratives less frequently than their younger counterparts. Kallel (2002: 177) also finds that *do*-support is favoured among younger speakers, particularly in negative declaratives.

The data from the *Salem Witchcraft Papers* appears to show the reverse effect. *Do*-support is used most frequently by the oldest subset of speakers (20%). The youngest and the middle-aged group of speakers show similar rates (10% and 8% respectively), suggesting that the proper split may be between those older than 50 and those younger than 50.

| Factors      | <i>do</i> -support |             | simple    |              | <b>Total</b> |              |
|--------------|--------------------|-------------|-----------|--------------|--------------|--------------|
|              | %                  | N           | %         | N            | %            | N            |
| Younger      | 10                 | 370         | 90        | 3249         | <b>28</b>    | <b>3619</b>  |
| Middle       | 8                  | 291         | 92        | 3138         | <b>26</b>    | <b>3429</b>  |
| Older        | 20                 | 551         | 80        | 2229         | <b>21</b>    | <b>2780</b>  |
| Unknown      | 19                 | 633         | 81        | 2676         | <b>25</b>    | <b>3309</b>  |
| <b>Total</b> | <b>14</b>          | <b>1845</b> | <b>86</b> | <b>11292</b> | <b>100</b>   | <b>13137</b> |

Table 93: Contribution of speaker age to the selection of *do*-support

Given that the literature points to a correlation between age and sentence type, further examination of this relationship is a fertile avenue for investigation. Table 94 below does in fact dovetail with the claims made by Warner (2005) and Kallel (2002), as negative declarative *do* is most frequently used by the youngest

subset of speakers (45%). Here, the middle-aged and older speakers pattern very similarly (33% and 32% respectively). Older speakers predominate in the use of affirmative interrogative *do* with a rate of 51%. Negative interrogatives show only minor variation between the three age groups. There are only a small number of negative imperative tokens, but young speakers appear to favour the use of *do*-support in this context (80%, but based on only 5 tokens for younger speakers). Taken together, these results appear to suggest that older speakers may be leading in the use of interrogative *do* (particularly in affirmatives), while younger and middle-aged speakers lag behind. Contrastively, younger speakers are leading in the use of *do*-support in negatives (particularly negative declaratives), while middle-aged and older speakers lag behind. The proportionally higher rate of affirmative declarative *do* used by older speakers (11%, compared to 6% for middle-aged speakers and 7% for younger speakers), a construction which historically was in decline (Ellegård, 1953), is as one would expect from an apparent-time window on this change in progress.

| Factors                   | Young              |     |        |      | Middle             |     |        |      | Old                |     |        |      | Total |      |
|---------------------------|--------------------|-----|--------|------|--------------------|-----|--------|------|--------------------|-----|--------|------|-------|------|
|                           | <i>do</i> -support |     | simple |      | <i>do</i> -support |     | simple |      | <i>do</i> -support |     | simple |      |       |      |
|                           | %                  | N   | %      | N    | %                  | N   | %      | N    | %                  | N   | %      | N    | %     | N    |
| Affirmative Declarative   | 7                  | 203 | 93     | 2811 | 6                  | 188 | 94     | 2808 | 11                 | 216 | 89     | 1778 | 81    | 8004 |
| Affirmative Interrogative | 17                 | 65  | 83     | 314  | 13                 | 25  | 87     | 170  | 51                 | 184 | 49     | 178  | 10    | 936  |
| Affirmative Imperative    | 0                  | 0   | 100    | 15   | 0                  | 0   | 100    | 20   | 2                  | 1   | 98     | 63   | 1     | 99   |
|                           |                    |     |        |      |                    |     |        |      |                    |     |        |      |       |      |
| Negative Declarative      | 45                 | 85  | 55     | 106  | 33                 | 60  | 67     | 124  | 32                 | 93  | 68     | 197  | 7     | 665  |
| Negative Interrogative    | 87                 | 13  | 13     | 2    | 94                 | 17  | 6      | 1    | 92                 | 54  | 8      | 5    | 1     | 92   |

|                     |          |            |           |             |          |            |           |             |          |            |           |             |            |             |
|---------------------|----------|------------|-----------|-------------|----------|------------|-----------|-------------|----------|------------|-----------|-------------|------------|-------------|
| Negative Imperative | 80       | 4          | 20        | 1           | 6        | 1          | 94        | 15          | 27       | 3          | 73        | 8           | <b>0.3</b> | <b>32</b>   |
| <b>Total</b>        | <b>4</b> | <b>370</b> | <b>33</b> | <b>3249</b> | <b>3</b> | <b>291</b> | <b>32</b> | <b>3138</b> | <b>6</b> | <b>551</b> | <b>23</b> | <b>2229</b> | <b>100</b> | <b>9828</b> |

Table 94: Contribution of speaker age and sentence type to the selection of *do*-support

Further cross-tabulation of age, sex, and sentence type reveals that, across age categories, female speakers lead in the use of declarative *do* (affirmative and negative). The only sentence type where male speakers use *do*-support more frequently is with affirmative interrogatives. This finding appears to suggest that male speakers may be leading the change in the decline of affirmative declarative *do*, while also leading the increase in (affirmative) interrogative *do*.

| Factors                   | Young Male |           | Young Female |            | Middle Male |            | Middle Female |           | Old Male |            | Old Female |            | Total      |              |
|---------------------------|------------|-----------|--------------|------------|-------------|------------|---------------|-----------|----------|------------|------------|------------|------------|--------------|
|                           | %          | N         | %            | N          | %           | N          | %             | N         | %        | N          | %          | N          | %          | N            |
| Affirmative Declarative   | 5          | 33/637    | 7            | 170/2377   | 6           | 135/2407   | 9             | 53/589    | 10       | 152/1471   | 12         | 64/523     | <b>79</b>  | <b>10408</b> |
| Affirmative Interrogative | 15         | 8/55      | 28           | 57/324     | 10          | 13/133     | 19            | 12/62     | 54       | 170/316    | 30         | 14/46      | <b>11</b>  | <b>1501</b>  |
| Negative Declarative      | 32         | 12/38     | 48           | 73/153     | 31          | 37/121     | 37            | 23/63     | 27       | 28/104     | 35         | 65/186     | <b>6</b>   | <b>794</b>   |
| Other <sup>24</sup>       | 60         | 3/5       | 47           | 14/30      | 26          | 12/46      | 75            | 6/8       | 47       | 55/118     | 19         | 3/16       | <b>3</b>   | <b>434</b>   |
| <b>Total</b>              | <b>0.4</b> | <b>56</b> | <b>2</b>     | <b>314</b> | <b>1</b>    | <b>197</b> | <b>1</b>      | <b>94</b> | <b>3</b> | <b>405</b> | <b>1</b>   | <b>146</b> | <b>100</b> | <b>13137</b> |

Table 95: Contribution of speaker age, speaker sex, and sentence type to the selection of *do*-support

#### 4.4.4.3 Birthplace

Claims have been made (Ellegård, 1953; Söderlund, 2017: 273) that *do*-support may be subject to regional variation. Rissanen (2003: 108) suggests that affirmative declarative *do* in particular may have been more common in early American English than in the British English of the 17<sup>th</sup> century. It should be noted, however, that this claim is necessarily based on written texts, with all its attendant limitations.

<sup>24</sup> Contains affirmative imperatives, negative imperatives, and negative interrogatives.

Other potential infelicities include frequency calculations based on rates of *do*-support per 10,000 words, an approach which, as mentioned previously, contravenes the Principle of Accountability (Labov, 1972), requiring the analyst to take into account contexts where *do*-support occurs as well as those where it does not, but potentially could have done so. Notwithstanding these caveats, in the *Salem Witchcraft Papers*, it can be seen that *do*-support is comparatively less frequent within the cohort of speakers who were born in Britain, with a rate of 10%. This has potential implications for the concept of ‘colonial lag’ (Marckwardt, 1958), according to which transplanted varieties (in this case American English) are more conservative linguistically than their source varieties (Hundt, 2009).

| Factors       | <i>do</i> -support |             | simple    |              | Total        |              |
|---------------|--------------------|-------------|-----------|--------------|--------------|--------------|
|               | %                  | N           | %         | N            | %            | N            |
| New England   | 14                 | 844         | 86        | 5341         | <b>47</b>    | <b>6185</b>  |
| Britain       | 10                 | 159         | 90        | 1471         | <b>12</b>    | <b>1630</b>  |
| South America | 12                 | 37          | 88        | 284          | <b>2</b>     | <b>321</b>   |
| Barbados      | 0                  | 0           | 100       | 4            | <b>0.003</b> | <b>4</b>     |
| Unknown       | 16                 | 805         | 84        | 4192         | <b>38</b>    | <b>4997</b>  |
| <b>Total</b>  | <b>14</b>          | <b>1845</b> | <b>86</b> | <b>11292</b> | <b>100</b>   | <b>13137</b> |

Table 96: Contribution of speaker birthplace to the selection of *do*-support

Cross-tabulating speaker birthplace with sentence type suggests that American English may have been leading in the increase in the use of *do*-support in negatives and interrogatives. American-born speakers use *do*-support more frequently than British-born speakers in affirmative interrogatives (33% to 24%), negative declaratives (39% to 27%), negative interrogatives (92% to 82%) and negative imperatives (36% to 12%). The rates for affirmative declarative *do* are almost identical between American-born and British-born speakers. As will be shown in Section 4.4.5, the difference in the use of *do*-support between American-born and British-born speakers is statistically significant.

| Factors | America | Britain | South America | Total |
|---------|---------|---------|---------------|-------|
|---------|---------|---------|---------------|-------|

|                           | <i>do</i> -support |            | simple    |             | <i>do</i> -support |            | simple    |             | <i>do</i> -support |           | simple   |            |            |             |
|---------------------------|--------------------|------------|-----------|-------------|--------------------|------------|-----------|-------------|--------------------|-----------|----------|------------|------------|-------------|
|                           | %                  | N          | %         | N           | %                  | N          | %         | N           | %                  | N         | %        | N          | %          | N           |
| Affirmative Declarative   | 8                  | 369        | 92        | 4536        | 7                  | 94         | 93        | 1274        | 5                  | 13        | 95       | 233        | <b>80</b>  | <b>6519</b> |
| Affirmative Interrogative | 33                 | 237        | 67        | 474         | 24                 | 15         | 76        | 47          | 8                  | 1         | 92       | 12         | <b>10</b>  | <b>786</b>  |
| Affirmative Imperative    | 2                  | 1          | 98        | 59          | 0                  | 0          | 100       | 28          | 0                  | 0         | 100      | 23         | <b>1</b>   | <b>111</b>  |
|                           |                    |            |           |             |                    |            |           |             |                    |           |          |            |            |             |
| Negative Declarative      | 39                 | 167        | 61        | 257         | 27                 | 39         | 73        | 105         | 57                 | 20        | 43       | 15         | <b>7</b>   | <b>603</b>  |
| Negative Interrogative    | 92                 | 65         | 8         | 6           | 82                 | 9          | 18        | 2           | 100                | 3         | 0        | 0          | <b>1</b>   | <b>85</b>   |
| Negative Imperative       | 36                 | 5          | 64        | 9           | 12                 | 2          | 88        | 15          | 0                  | 0         | 100      | 1          | <b>0.4</b> | <b>32</b>   |
| <b>Total</b>              | <b>10</b>          | <b>844</b> | <b>66</b> | <b>5341</b> | <b>2</b>           | <b>159</b> | <b>18</b> | <b>1471</b> | <b>0.4</b>         | <b>37</b> | <b>3</b> | <b>284</b> | <b>100</b> | <b>8136</b> |

Table 97: Contribution of speaker birthplace and sentence type to the selection of *do*-support

#### 4.4.4.4 Social Status

*Do*-support has not been directly associated with social status in the extant literature, although it has been correlated with prestige and formality (Cuesta, 2016: 133; Nurmi, 2018: 127; Pérez, 1997: 39; Warner, 2017: 229-230; Wischer, 2008: 142-143). As Table 98 demonstrates, there is little differentiation between the social groups operationalized in this study in their use of *do*-support. The two lower status groups (C and D) use *do*-support slightly less frequently, but this is a minor effect.

| Factors | <i>do</i> -support |     | simple |      | <b>Total</b> |             |
|---------|--------------------|-----|--------|------|--------------|-------------|
|         | %                  | N   | %      | N    | %            | N           |
| Group A | 16                 | 360 | 84     | 1890 | <b>17</b>    | <b>2250</b> |

|              |           |             |           |              |            |              |
|--------------|-----------|-------------|-----------|--------------|------------|--------------|
| Group B      | 14        | 344         | 86        | 2121         | <b>19</b>  | <b>2465</b>  |
| Group C      | 11        | 225         | 89        | 1769         | <b>15</b>  | <b>1994</b>  |
| Group D      | 11        | 37          | 89        | 310          | <b>3</b>   | <b>347</b>   |
| Unknown      | 14        | 879         | 86        | 5202         | <b>46</b>  | <b>6081</b>  |
| <b>Total</b> | <b>14</b> | <b>1845</b> | <b>86</b> | <b>11292</b> | <b>100</b> | <b>13137</b> |

Table 98: Contribution of speaker social status to the selection of *do*-support

Cross-tabulation of speaker social status with sentence type shows an interesting effect. Negative interrogatives are only used to any appreciable extent by members of Group A, which may reflect the fact that many of the speakers in Group A are court officials, who are interrogating the accused as part of the examinations (and therefore frequently ask questions). There is an apparent social status effect on the use of affirmative interrogatives and negative declaratives. The highest status speakers (Group A) use *do*-support notably more frequently in affirmative interrogatives (55%) than any other speaker group, which again may be related to their role in the trial. There is a clear effect of social status on the use of *do*-support in negative declaratives, with the rate of *do*-support consistently increasing as speaker social status decreases. Therefore Group A uses *do*-support in negative declaratives the least frequently (25%), followed by Group B (35%), then Group C (41%). Group D uses *do*-support in negative declaratives the most frequently (56%), though this group is made up of only three speakers (Tituba, Candy, and John Indian), and produces only 36 negative declaratives collectively.

| Factors                   | Group A            |     |        |      | Group B            |     |        |      | Group C            |     |        |      | Group D            |    |        |     | Total     |             |
|---------------------------|--------------------|-----|--------|------|--------------------|-----|--------|------|--------------------|-----|--------|------|--------------------|----|--------|-----|-----------|-------------|
|                           | <i>do</i> -support |     | simple |      | <i>do</i> -support |     | simple |      | <i>do</i> -support |     | simple |      | <i>do</i> -support |    | simple |     |           |             |
|                           | %                  | N   | %      | N    | %                  | N   | %      | N    | %                  | N   | %      | N    | %                  | N  | %      | N   | %         | N           |
| Affirmative Declarative   | 7                  | 117 | 93     | 1605 | 10                 | 200 | 90     | 1739 | 7                  | 121 | 93     | 1518 | 5                  | 13 | 95     | 254 | <b>79</b> | <b>5567</b> |
| Affirmative Interrogative | 55                 | 173 | 45     | 139  | 17                 | 45  | 83     | 218  | 20                 | 36  | 80     | 147  | 6                  | 1  | 94     | 16  | <b>11</b> | <b>775</b>  |
| Affirmative Imperative    | 2                  | 1   | 98     | 62   | 0                  | 0   | 100    | 11   | 0                  | 0   | 100    | 12   | 0                  | 0  | 100    | 23  | <b>2</b>  | <b>109</b>  |
|                           |                    |     |        |      |                    |     |        |      |                    |     |        |      |                    |    |        |     |           |             |

|                        |          |            |           |             |          |            |           |             |          |            |           |             |          |           |          |            |            |             |
|------------------------|----------|------------|-----------|-------------|----------|------------|-----------|-------------|----------|------------|-----------|-------------|----------|-----------|----------|------------|------------|-------------|
| Negative Declarative   | 25       | 18         | 75        | 53          | 35       | 82         | 65        | 151         | 41       | 63         | 59        | 91          | 56       | 20        | 44       | 16         | 7          | 494         |
| Negative Interrogative | 91       | 50         | 9         | 5           | 88       | 14         | 13        | 2           | 100      | 4          | 0         | 0           | 100      | 3         | 0        | 0          | 1          | 78          |
| Negative Imperative    | 4        | 1          | 96        | 26          | 100      | 3          | 0         | 0           | 50       | 1          | 50        | 1           | 0        | 0         | 100      | 1          | 0.4        | 33          |
| <b>Total</b>           | <b>5</b> | <b>360</b> | <b>27</b> | <b>1890</b> | <b>5</b> | <b>344</b> | <b>30</b> | <b>2121</b> | <b>3</b> | <b>225</b> | <b>25</b> | <b>1769</b> | <b>1</b> | <b>37</b> | <b>4</b> | <b>310</b> | <b>100</b> | <b>7056</b> |

Table 99: Contribution of speaker social status and sentence type to the selection of *do*-support

#### 4.4.4.5 Document Type

In the 16<sup>th</sup> century, the use of *do*-support has been claimed to be associated with formal contexts (Cuesta, 2016; Warner, 2017). Wischer (2008) observes that *do*-support may be more common in formal contexts in the early 16<sup>th</sup> century, but that it becomes increasingly associated with informal contexts. Pérez (1997) finds *do*-support to be more frequent in texts that more closely approximate speech, and rarer in those belonging to more formal written genres.

Using text type as a diagnostic of formality and proximity to speech, the data from the *Salem Witchcraft Papers* suggests that the association of *do*-support with less formal, spoken contexts has almost disappeared by the end of the 17<sup>th</sup> century. Examinations, which most closely approximate spoken language, show comparatively higher rates of *do*-support, at 18%. Testimonies (encompassing testimonies and witness depositions) and personal documents (such as letters and petitions) show identical rates of *do*-support (11%).

| Factors            | <i>do</i> -support |             | simple    |              | Total      |              |
|--------------------|--------------------|-------------|-----------|--------------|------------|--------------|
|                    | %                  | N           | %         | N            | %          | N            |
| Examinations       | 18                 | 1021        | 82        | 4748         | 44         | 5769         |
| Testimonies        | 11                 | 592         | 89        | 4965         | 42         | 5557         |
| Personal Documents | 11                 | 112         | 89        | 889          | 8          | 1001         |
| Official Documents | 15                 | 120         | 85        | 690          | 6          | 810          |
| <b>Total</b>       | <b>14</b>          | <b>1845</b> | <b>86</b> | <b>11292</b> | <b>100</b> | <b>13137</b> |



Table 100: Contribution of document type to the selection of *do*-support

It is interesting to note that the least formal (examinations) and most formal (official documents) text types show the highest rates for *do*-support. Cross-tabulating text type with sentence type reveals an interesting pattern. Affirmative declarative *do* is at its most frequent in official documents (15%), and its least frequent in examinations (5%). Negative declarative *do* is also more frequent in official documents (67%). Examinations otherwise show the highest rates of *do*-support in all other sentence types, as would be expected if *do*-support has its origins in the spoken language, and suggests a relationship between *do*-support and less formal contexts. The surprising and relatively high rate of *do*-support in affirmative declaratives in official documents invites further inquiry. One possible explanation is that the decline of affirmative declarative *do*, and its rise in other sentence types, is a development that is fuelled by the spoken as opposed to the written language, which is in accordance with the generally held notion that change typically begins in speech (Labov, 1972). If this is the case, it might be expected that more conservative text types will lag behind, and show a higher rate of *do*-support in affirmative declaratives. The data from the *Salem Witchcraft Papers* appear to support this hypothesis. Conversely, if the rise in the use of *do*-support in interrogatives has its origin in the spoken language, then it might be expected that less formal, more speech-like texts (e.g. examinations) would show higher rates of *do*-support in interrogatives, and more conservative, formal texts (e.g. official documents) would show lower rates for interrogatives. This appears to be the case, though it should be noted that interrogatives are generally rare in official documents.

| Factors                   | Examinations       |     |        |      | Testimonies        |     |        |      | Personal Documents |    |        |     | Official Documents |    |        |     | Total |       |
|---------------------------|--------------------|-----|--------|------|--------------------|-----|--------|------|--------------------|----|--------|-----|--------------------|----|--------|-----|-------|-------|
|                           | <i>do</i> -support |     | simple |      | <i>do</i> -support |     | simple |      | <i>do</i> -support |    | simple |     | <i>do</i> -support |    | simple |     |       |       |
|                           | %                  | N   | %      | N    | %                  | N   | %      | N    | %                  | N  | %      | N   | %                  | N  | %      | N   | %     | N     |
| Affirmative Declarative   | 5                  | 187 | 95     | 3959 | 9                  | 422 | 91     | 4315 | 9                  | 75 | 91     | 775 | 15                 | 99 | 85     | 576 | 79    | 10408 |
| Affirmative Interrogative | 59                 | 491 | 41     | 336  | 14                 | 78  | 86     | 479  | 16                 | 10 | 84     | 52  | 20                 | 11 | 80     | 44  | 11    | 1501  |
| Affirmative Imperative    | 1                  | 1   | 99     | 153  | 0                  | 0   | 100    | 14   | 0                  | 0  | 100    | 5   | 0                  | 0  | 100    | 26  | 2     | 199   |

|                        |          |             |           |             |          |            |           |             |          |            |          |            |          |            |          |            |            |              |
|------------------------|----------|-------------|-----------|-------------|----------|------------|-----------|-------------|----------|------------|----------|------------|----------|------------|----------|------------|------------|--------------|
| Negative Declarative   | 40       | 190         | 60        | 285         | 32       | 73         | 68        | 154         | 27       | 21         | 73       | 56         | 67       | 10         | 33       | 5          | 6          | 794          |
| Negative Interrogative | 92       | 145         | 8         | 12          | 94       | 17         | 6         | 1           | 83       | 5          | 17       | 1          | 0        | 0          | 0        | 0          | 1          | 181          |
| Negative Imperative    | 70       | 7           | 30        | 3           | 50       | 2          | 50        | 2           | 100      | 1          | 0        | 0          | 0        | 0          | 100      | 39         | 0.4        | 54           |
| <b>Total</b>           | <b>8</b> | <b>1021</b> | <b>36</b> | <b>4748</b> | <b>5</b> | <b>592</b> | <b>38</b> | <b>4965</b> | <b>1</b> | <b>112</b> | <b>7</b> | <b>889</b> | <b>1</b> | <b>120</b> | <b>5</b> | <b>690</b> | <b>100</b> | <b>13137</b> |

Table 101: Contribution of document type and sentence type to the selection of *do*-support

#### 4.4.5 Statistical Analysis

Tables 102 and 103 below present the results of a mixed-effects regression analysis of the contribution of linguistic and social factor groups respectively to the use of *do*-support using Rbrul (Johnson, 2009). To control for the possibility of individual effects (understood here to refer to the attribution of specific tokens to particular individuals), building on suggestions that the use of *do*-support may be idiosyncratic (Jones & Tagliamonte, 2004: 102; Nurmi, 2018: 126; Pérez, 1997: 39), individual was run as a random effect (Johnson, 2009). It is important to reiterate the caveat that caution should be exercised in equating individual tokens with verbatim utterances, given the omnipresent issue of scribal intervention (as discussed in more detail in Section 2.3). Invariable individuals (who either only used *do*-support, or never used it) were excluded from the model.

|                                                     |           |            |   |   |
|-----------------------------------------------------|-----------|------------|---|---|
| Input probability                                   | 0.363     |            |   |   |
| Log likelihood                                      | -3458.065 |            |   |   |
| R2 total                                            | 0.445     |            |   |   |
| N                                                   | 11021     |            |   |   |
|                                                     |           |            |   |   |
| <b>Polarity/Sentence Type</b><br><b>(p &lt; ~0)</b> | Log Odds  | Centred FW | % | N |

|                                                     |        |       |    |      |
|-----------------------------------------------------|--------|-------|----|------|
| Negative Interrogative                              | 2.507  | 0.925 | 92 | 176  |
| Affirmative Interrogative                           | 0.337  | 0.583 | 43 | 1364 |
| Negative Declarative                                | -0.250 | 0.438 | 40 | 671  |
| Affirmative Declarative                             | -2.594 | 0.07  | 9  | 8810 |
| <b>Verb Morphology</b><br><b>(p &lt; 1.07e-105)</b> |        |       |    |      |
| Weak                                                | 0.401  | 0.599 | 18 | 4438 |
| Strong                                              | -0.401 | 0.401 | 14 | 6485 |
| <b>Clause Type</b><br><b>(p &lt; 5.55e-85)</b>      |        |       |    |      |
| Main                                                | 0.992  | 0.730 | 18 | 7255 |
| Other Subordinate                                   | 0.611  | 0.648 | 15 | 2688 |
| Relative                                            | -1.603 | 0.168 | 8  | 1078 |
| <b>Adverb</b><br><b>(p &lt; 5.06e-10)</b>           |        |       |    |      |
| Adverb                                              | 0.253  | 0.563 | 17 | 2095 |
| No Adverb                                           | -0.253 | 0.437 | 16 | 8926 |
| <b>Tense</b><br><b>(n.s.)</b>                       |        |       |    |      |
| Present                                             | 0.031  | 0.508 | 24 | 2404 |
| Past                                                | -0.031 | 0.492 | 14 | 8617 |

Table 102: Rbrul mixed-effects logistic regression analysis of the contribution of linguistic factors to the selection of do-support<sup>25</sup>

<sup>25</sup> Tokens that were too infrequent (e.g. imperatives; semi-weak verb morphology; habitual aspect; adverb position) or where the context was ambiguous or unclear (e.g. tense) are excluded here. Factor groups like aspect or adverb position were not analysed due to insufficient token counts for one or more factors.

The results in Table 102 above show that, of the factor groups tested, only tense was not a statistically significant predictor of *do*-support. Even so, present tense contexts slightly favour *do*-support, while past tense contexts slightly disfavour.

Sentence type exerts a notable effect on *do*-support, and the direction of effect dovetails with findings reported in Ellegård (1953: 166), with interrogatives exerting a stronger effect on the use of *do*-support than declaratives, and negatives exerting a stronger effect than their affirmative counterparts. The major divergences from Ellegård's (1953) findings are the comparatively higher rate of affirmative declarative *do* here (9% compared to Ellegård's 1.8% for the same time period), and the lower rate for affirmative interrogatives (43% compared to 79.2% for Ellegård). The latter difference may be attributable to genre effects, as Ellegård's findings were based on prose and literary texts.

*Do*-support has previously been associated with strong verbal morphology (Bækken, 2002: 326; Ellegård, 1953). As Table 102 shows, such an association is unfounded in the data, as weak verb morphology favours *do*-support (though only slightly).

As has been mentioned (e.g. Section 4.4.2.4), there are numerous conflicting claims in the literature on the contribution of clause type to the selection of *do*-support. The results reported in Table 102 show main clauses and (non-relative) subordinate clauses as favouring *do*-support, which is in line with claims by Stein (1990), Ihalainen (1991), and Jones and Tagliamonte (2004: 113). The disfavouring effect of relative clauses on the use of *do*-support is at odds with Rissanen's (1985: 173) claim that *do*-support is favoured in relative clauses.

The presence of an adverb is a statistically significant predictor for the use of *do*-support. This supports claims made by Bækken (2002), Ellegård (1953), Nurmi (2018), and Partridge (1948).

Table 103 below presents the contribution of social factor groups to the use of *do*-support. Speakers for whom sociobiographical information was absent from the historical record were excluded from this analysis, hence the difference in Total Ns in Tables 102 and 103.

|                                                          |           |            |    |      |
|----------------------------------------------------------|-----------|------------|----|------|
| Input probability                                        | 0.104     |            |    |      |
| Log likelihood                                           | -2554.139 |            |    |      |
| R2 total                                                 | 0.155     |            |    |      |
| N                                                        | 7083      |            |    |      |
|                                                          |           |            |    |      |
| <b>Age</b><br><b>(p &lt; 2.12e-51)</b>                   | Log Odds  | Centred FW | %  | N    |
| Older                                                    | 0.723     | 0.673      | 23 | 2193 |
| Younger                                                  | -0.232    | 0.442      | 11 | 2744 |
| Middle                                                   | -0.491    | 0.38       | 8  | 2146 |
| <b>Sex</b><br><b>(p &lt; 0.00159)</b>                    |           |            |    |      |
| Female                                                   | 0.257     | 0.564      | 14 | 3299 |
| Male                                                     | -0.257    | 0.436      | 14 | 3784 |
| <b>Birthplace<sup>26</sup></b><br><b>p &lt; 0.00451)</b> |           |            |    |      |
| America                                                  | 0.319     | 0.579      | 15 | 5606 |
| Britain                                                  | -0.319    | 0.421      | 9  | 1477 |
| <b>Document type</b><br><b>(p &lt; 0.0242)</b>           |           |            |    |      |
| Examinations                                             | 0.180     | 0.545      | 15 | 3421 |
| Testimonies                                              | 0.040     | 0.51       | 13 | 2642 |
| Official Documents                                       | -0.063    | 0.484      | 12 | 347  |

<sup>26</sup> Speakers from outside of America or Britain were excluded from this analysis as they accounted for only a small number of tokens.

|                                                  |        |       |    |      |
|--------------------------------------------------|--------|-------|----|------|
| Personal Documents                               | -0.157 | 0.461 | 11 | 673  |
| <b>Role in Trial</b><br><b>(p &lt; 5.47e-13)</b> |        |       |    |      |
| Official                                         | 0.356  | 0.588 | 19 | 1607 |
| Accused                                          | 0.154  | 0.539 | 16 | 1746 |
| Witness                                          | -0.208 | 0.448 | 11 | 1759 |
| Accuser                                          | -0.303 | 0.425 | 11 | 1971 |

Table 103: Rbrul mixed-effects logistic regression analysis of the contribution of social factors to the selection of *do*-support<sup>27</sup>

Age exerts an influence on the use of *do*-support, which is favoured among the oldest speakers. Younger and middle-aged speakers disfavour *do*-support. This finding is at odds with previous research, which reports that *do*-support is favoured among younger speakers (Kallel, 2002: 177; Warner, 2005: 268-270). However, it should be noted that these claims are based on negative declaratives only. As mentioned in Section 4.4.4.2, the disfavouring of *do*-support among younger speakers (and relative favouring of *do*-support among older speakers) may derive from the higher rates of affirmative declarative *do* in older speakers.

*Do*-support is preferred if the speaker is female, and dispreferred if the speaker is male. This is in line with claims advanced by Nurmi (2018: 129), who associates the rise in *do*-support with female speakers, but claims that sex effects have ceased to influence the use of *do*-support in the 17<sup>th</sup> century. The statistical significance of sex in this model suggests that sex-related effects may persist for longer than is claimed by Nurmi (2018).

Speakers who were born in New England were statistically more likely to use *do*-support than those born in Britain. This lends some support to Rissanen's (2003: 108) claim that *do*-support is more frequent among speakers from America. However, contemporaneous data from speakers in Britain would be needed to strengthen this claim.

<sup>27</sup> Social status was excluded from this analysis.

*Do*-support was favoured in examinations and testimonies, the two text types that most closely approximate spoken language (with those two text types construed here as a measure of informality), and disfavoured in official documents and personal documents. In the literature, it has been hypothesised that while *do*-support was favoured in formal contexts in the 16<sup>th</sup> century (Cuesta, 2016; Warner, 2017), in the 17<sup>th</sup> century it developed an increasing association with informal contexts (Ellegård, 1953; Pérez, 1997; Wischer, 2008). The results in Table 103 seem to at least support the 17<sup>th</sup> century association with informal contexts.

The speaker's role in the trial was also a significant predictor of *do*-support usage. *Do*-support was used most frequently by court officials and the accused, and least frequently by witnesses and accusers. The ordering of effects for this factor group could be related to the frequency of sentence types in each context. As has been demonstrated, interrogatives and negative declaratives appear with *do*-support more frequently than affirmative declaratives. In the *Salem Witchcraft Papers*, the accusers and witnesses were most likely to use affirmative declaratives, as they are recounting past events. It would not be unexpected for these speakers to use less *do*-support. Court officials frequently used interrogatives, as they are questioning the accused (particularly during examinations), while the accused made frequent use of negative declaratives as they denied the accusations levied against them. These speakers might therefore be expected to use *do*-support more frequently, which appears to be the case.

It is worth noting that the findings in Table 103 do not capture the important interaction between social factors and sentence type (as discussed in more detail in Section 4.4.4).

## 4.5 Discussion

The examination of the use of *do*-support in the *Salem Witchcraft Papers* has provided critical insights into syntactic variation in 17<sup>th</sup> century colonial English as well as Early Modern English more generally. The methodological strengths of the variationist framework, coupled with the large amount of data brought to bear on the variable context and the use of statistical analysis, have shed light on numerous facets of the variable system underpinning *do*-support in the late 17<sup>th</sup> century. The analysis targeted a variety of factor groups abstracted from the literature, which were operationalised to explore the

distribution of *do*-support across linguistic contexts, and to identify the social embedding of the phenomenon at the end of the 17<sup>th</sup> century. We are now in a position to explore the contribution of our major findings to the primary research objectives detailed earlier.

#### 4.5.1 Objective (i): to reconstruct the variable system of the target linguistic phenomenon as it exists in contemporaneous speech-related data

The current study has used distributional and multivariate analyses to identify the internal linguistic and grammar-external factors influencing the use of *do*-support in 17<sup>th</sup> century English. This subsection will summarise the major findings arising from the analysis.

Sentence type is the single most important predictor for the use of *do*-support, and exerts an influence over every other factor group tested. In each case, cross-tabulation of the relevant factor group with sentence type revealed important results that would have otherwise remained hidden. This is congruent with the importance attached to sentence type in the literature (Ellegård, 1953: 161; Kroch, 1989; Ogura, 1993; Oudesluijs et al., 2022; Söderlund, 2017: 260). Negative interrogatives are clearly the preferred niche for *do*-support, and generally favour it across contexts.

Tense, aspect, and verb morphology are three factor groups that are possibly interconnected. Denison (1993: 283) suggests that *do*-support may have been used as a tense marker to reduce ambiguity in verbs that do not show overt morphological tense differentiation in the past tense (i.e. *put*, *hurt*, *cut*). If this is the case, *do*-support would be favoured with verbs with strong morphology (which includes the verbs mentioned above), as is claimed by Bækken (2002: 326), and in the past tense. However, the results reported here do not support this, as strong verbs actually disfavour *do*-support. Although not statistically significant, there are tense effects present in the data, as *do*-support in interrogatives is found to be slightly more frequent in past tense contexts than in the present tense. If *do*-support was used to reduce ambiguity in the way claimed by Denison (1993: 283), there is no plausible reason that this would operate in interrogatives only.

In habitual past tense contexts, *do*-support is favoured in affirmative declaratives. Interestingly, habitual *do*-support is disfavoured in negative interrogatives, which are otherwise the preferred context for *do*-



support. Strength of verb morphology exhibits only a slight effect, with weak verbs favouring *do*-support in affirmative declaratives.

The presence of an adverb is a statistically significant predictor of *do*-support, favouring the variant in positive contexts (and consequently disfavouring it in negative contexts). The position of the adverb also plays a role – *do*-support in interrogatives is most frequent with a pre-verbal adverb, while negative declaratives favour *do*-support with a post-verbal adverb. *Do*-support in affirmative declaratives is more common when there is both a pre-verbal and a post-verbal adverb. The interrelationship of sentence type and adverb placement is a novel finding that has not previously been reported in the pertinent literature, to the best of my knowledge. It is also notable that where *do*-support appears in conjunction with an adverb, the adverb is most likely to intervene between *do* and the main verb.

Detailed investigation of the role played by specific verb collocations, based on the oft-cited *know*-group of verbs, which are claimed to resist the use of *do*-support, proved to be particularly instructive (Engblom, 1938; Ellegård, 1953; Nurmi, 2018; Warner, 2017). The majority of verbs in the data are very low-frequency, appearing only once or twice, and so analysis was restricted to the high-frequency verbs (those appearing 100+ times). This included verbs in the *know*-group (such as *know*, *come*, *see*, *go*). Among the high-frequency verbs, semantics effects were apparent: *do*-support was favoured in interrogatives with verbs expressing mental processes (e.g. *think*, *know*, *see*). Verbs expressing speech acts (e.g. *tell*, *ask*, *answer*, *cry*) disfavour *do*-support in affirmative declaratives, while highly emotive verbs like *afflict*, *hurt*, and *torment* are favoured in this context. An additional effect is tied to the etymological origins of particular verb forms, with Latinate verbs favouring *do*-support in declaratives and disfavouring it in interrogatives.

Investigation into the social embedding of *do*-support was equally revelatory. It is clear from the results that *do*-support is subject to social conditioning in late seventeenth-century American English, and the use of *do*-support in different sentence types by different social groups may provide some indication as to the origin and direction of change in the use of *do*-support. Age, sex, birthplace, role in trial and document type were all statistically significant predictors in the use of *do*-support.

*Do*-support is found to be subject to sex effects when cross-tabulated with sentence type. Male speakers used more affirmative interrogative *do*, while female speakers used more negative declarative *do*. Rates for *do*-support in affirmative declaratives and negative interrogatives show little differentiation between male and female speakers. Overall, female speakers were more likely to use *do*-support than male speakers.

The results for age point to *do*-support as a phenomenon undergoing change at the end of the 17<sup>th</sup> century. *Do*-support in affirmative declaratives is used most frequently by older speakers (and least frequently by the youngest speakers), as would be expected if that context is in decline. The youngest speaker group uses *do*-support in negative declaratives most frequently, suggesting that it is a more recent innovation. The association of *do*-support with negative interrogatives appears to already be well-established by 1692, as the rate is consistently high across age groups.

There is some divergence in the use of *do*-support between speakers born in America and speakers born in Britain, with the former using more *do*-support in interrogatives and negative declaratives. As Table 103 indicates, this difference is statistically significant.

Speaker social status was found to exert a slight effect on the use of *do*-support, with the highest status speakers using more *do*-support in affirmative interrogatives than other speaker groups. There was also a clear social status continuum in the use of negative declarative *do*, with the highest status speakers using it the least, and the lowest status speakers using it the most.

Statistical analysis reveals that *do*-support is favoured in examinations and testimonies, and disfavoured in official documents and personal documents. It was hypothesised that *do*-support would operate along a spectrum of formality, with the least formal (most speech-like) texts favouring, and the most formal texts disfavouring. The ranking of examinations and testimonies supports this, but the higher rate of *do*-support in official documents than personal documents is at odds with this idea. Cross-tabulation with sentence type reveals that formal contexts show the highest rate of *do*-support in affirmative declaratives, while the informal text types show very low rates of affirmative declarative *do*, but predominate in all other contexts (negative declaratives, and interrogatives). This finding is congruent with change in the use of *do*-support

originating in the spoken language. While not direct representations of speech, trial examinations are the closest historical analogue of spoken language available to the researcher (Kytö & Walker, 2003), and thus the working hypothesis is that, of the documents in the *Salem Witchcraft Papers*, examinations most closely represent the spoken language. Official documents occupy the opposite end of the spectrum, being more formalised documents that have no connection to an original speech act. The higher use of affirmative declarative *do* in more formal (and less speech-like) documents could be a reflection of the linguistically more conservative nature of these documents, while examinations (which more closely represent the spoken language) demonstrate lower levels of affirmative declarative *do*, but higher levels of *do*-support in negatives and interrogatives, which is in line with changes in the distribution of *do*-support reported for this period (Ellegård, 1953).

The above discussion illustrates the complexity of the variable system under investigation, and demonstrates that there are many factors at play in the distribution of *do*-support in 17<sup>th</sup> century English, both linguistic and social. Targeting a larger number of factor groups generates a fuller picture of the variable system in question. Discussion will now turn to the relationship between the findings reported here and claims in the established literature.

4.5.2 Objective (ii): to refute, confirm, or refine existing claims in the literature regarding the target phenomenon

#### 4.5.2.1 *Confirmation of previous claims*

Of the claims made in the literature, only one may be tentatively confirmed by these results. Jones and Tagliamonte (2004) claim that *do*-support is strongly associated with the habitual in the past tense (based on affirmative declaratives only). The results reported here, whereby *do*-support in habitual contexts is slightly more common in affirmative declaratives, appear to support this claim. It was not, however, possible to subject this claim to rigorous statistical analysis as habitual contexts were too infrequent in the data.

#### 4.5.2.2 *Refinement of previous claims*

The application of the variationist framework has enabled this research to refine previous treatments of *do*-support in Early Modern English by addressing methodological confounds and infelicities in earlier treatments of this variable. The Principle of Accountability is a key tenet of the

variationist framework established by Labov (1972), and one that pays methodological dividends. Extracting all instances where the target feature, *do*-support, could have been used, even if it was not, enables the analyst to contextualize the target feature proportionally in relation to the larger variable system to which it belongs. This has clear advantages over normalising frequencies of a target variable (e.g. Nurmi, 2018), which may be suitable for large-scale corpus-based work, but is methodologically insufficient for sociolinguistic research. The seminal study on *do*-support by Ellegård (1953), which forms the basis for a majority of subsequent work on the variable, estimates the amount of *do*-support (and the simple form) using a sample of each text studied. Such an approach is predicated on the assumption that a sample from any text is likely to be representative of the remainder of the text. As Walker (2010: 74) points out, this is a potentially confounding issue, because there is a tendency for linguistic variables to cluster together. Evidence for this clustering with *do*-support can be seen in Rissanen (1985: 168-169), who observes that *do*-support is concentrated in certain passages or paragraphs that are “emphatic or emotionally colored” (Rissanen, 1985: 168), although it is not clear how exactly “emphatic” or “emotional” colouring is quantified in that study.

Another methodological advantage of the current study is the cross-tabulation of sentence type with all other factor groups, which has proved indispensable in illuminating otherwise hidden patterns in the data, and, crucially, addressing inconsistent claims in the literature. Every factor group that was tested showed an interaction with sentence type in the distribution of *do*-support. The value of considering the data through the lens of sentence type cannot be overstated. This important methodological point also serves to underscore what may be missed or overlooked in previous research by restricting the analytical focus to certain sentence types only, such as affirmative declaratives (Bækken, 2002; Ihalainen, 1991; Rissanen, 1985; Wischer, 2008), and negatives (Díaz-Vera, 1999; Pérez, 1997; Warner, 2005).

Considering a wider range of possible contexts for *do*-support serves to address some of the inconsistent claims in the literature. This is particularly pertinent in the case of the effects of clause type and adverb effects. For example, *do*-support has separately been associated with main clauses (Stein, 1990: 32-34), subordinate clauses (Ihalainen, 1991; Jones & Tagliamonte, 2004; Samuels, 1972), and specifically with relative clauses (Nevalainen, 1991; Rissanen, 1985). It is not clear that Ihalainen (1991) considered relative clauses separately from other subordinate clauses, while Jones and Tagliamonte (2004: 121) found

no difference between relative clauses and other subordinate clauses in the use of *do*-support, and so analysed them as a single group.

Cross-tabulation of clause type with sentence type shows that affirmative declarative *do* is slightly more frequent in subordinate clauses, in line with claims made by Ihalainen (1991) and Jones and Tagliamonte (2004), who both exclusively examined affirmative declaratives. Consideration of other contexts like negatives and interrogatives shows that the relationship between *do*-support and (non-relative) subordinate clauses is restricted to affirmative declaratives. *Do*-support is favoured in main clauses if it occurs in an affirmative interrogative context.

It is also possible to refine claims made about the relationship between adverbs and *do*-support. Previous research has associated the presence of an adverb with *do*-support, particularly in affirmative declaratives (Bækken, 2002: 328; Ellegård, 1953; Nurmi, 2018: 122; Stein, 1990: 358). The results reported here suggest that it is affirmatives more generally that favour the use of *do*-support with an adverb. There are abundant claims that a pre-verbal adverb favours *do*-support (Ihalainen, 1976; Pérez, 1997; Wischer, 2008); however, in each case post-verbal adverbs were not considered. The data from the *Salem Witchcraft Papers* suggests that pre-verbal adverbs favour *do*-support in interrogatives only, while negative declaratives favour *do*-support if there is a post-verbal adverb. As noted previously, the presence of adverbs in both pre- and post-verbal positions in the same sentence is favourable to *do*-support in affirmative declaratives. These findings demonstrate the value in also considering post-verbal adverbs and cases with multiple adverbs, as these factors – which have not been examined previously – do appear to have an influence on the use of *do*-support.

The present study attempted to undertake a more detailed and extensive investigation into the effect of lexical collocations on *do*-support. This was based on the oft-cited *know*-group of verbs which are claimed to resist the use of *do*-support (Ellegård, 1953; Engblom, 1938; Nurmi, 2018; Warner, 2017). High-frequency verbs in the data included many of those in the *know*-group (such as *know*, *come*, *see*, *go*). Among the other high-frequency verbs, semantic effects were apparent – *do*-support was favoured in interrogatives with verbs expressing mental processes (e.g. *think*, *know*, *see*). Verbs expressing speech acts (e.g. *tell*, *ask*, *answer*, *cry*) disfavour *do*-support in affirmative declaratives, while highly emotive verbs like *afflict*, *hurt*, and *torment* are favoured in this context.

Language of origin (either Latinate or Germanic) was also tested as a possible factor group, with the notable effect that Latinate verbs tended to favour *do*-support in declaratives, and disfavour it in interrogatives. This supports claims made by Denison (1993: 283) and Samuels (1972: 174).

With regards to social factor groups, claims about sex effects (e.g. Nurmi, 2018) are centred on female speakers using *do*-support more frequently than male speakers (and therefore leading the change). However, cross-tabulation with sentence type suggests that the situation is more complex than this. While female speakers predominate in the use of *do*-support in negative declaratives, male speakers are leading in the use of *do*-support in affirmative interrogatives.

Cross-tabulation with sentence type is also instructive when looking at the effect of age, where previous claims associating *do*-support in negatives with younger speakers may be refined (Kallel, 2002: 177; Warner, 2005: 268-269). It is important to note that Kallel's (2002) results were based on letters from the mid-16<sup>th</sup> century, and only included declaratives (as interrogatives were absent from their data), while Warner's (2005) data is from plays and prose from 1500-1710, and focused exclusively on negatives. The results reported here offer a more fine-grained and comprehensive picture. The youngest speaker group use more negative declarative *do* than other age groups, while older speakers use the highest rate of affirmative declarative *do*. This pattern is in line with a change in progress, and supports claims (e.g. by Ellegård, 1953) that *do*-support is declining in affirmative declaratives and increasing in other contexts. What is perhaps most interesting when looking at age is that *do*-support in negative interrogatives is stable across age groups.

#### 4.5.2.3 *Refutation of previous claims*

The results reported here call certain previous claims into question. Chief among these is the claim made by Kroch (1989) that *do*-support in affirmative declaratives had effectively disappeared from the language by 1700. This claim is based on the data used by Ellegård (1953), who reports a rate of 1.8% *do*-support in affirmative declaratives. Recall also that Ellegård's (1953) data was sub-optimal in the sense that they were based on conventional *written* genres, which tend to be more conservative than their speech-related counterparts. The current results suggest that the decline in the use of *do*-support in affirmative declaratives by 1700 is in need of some refinement, in light of the aggregated rate of 8% affirmative

declarative *do* in the *Salem Witchcraft Papers* (1692-1693), and somewhat higher rates in specific social groups (e.g. 11% in older speakers).

The literature on the relationship between tense and *do*-support claims that *do*-support is favoured in the past tense (Bækken, 2002: 325-326; Söderlund, 2017: 261). It should be noted that Bækken (2002) only looks at affirmative declaratives. Söderlund (2017) includes more contexts that could host *do*-support (negatives and interrogatives), and finds affirmative declarative *do* to be primarily associated with the past tense. The results reported here suggest that the association of *do*-support with the past tense is only the case in interrogatives, as declaratives show higher rates of *do*-support in the present tense than the past tense. This is in line with findings reported in Oudesluijs et al. (2022: 17) and Wischer (2008: 146), who also only consider affirmative declaratives.

The results from this study demonstrate the value of the application of an empirically accountable quantitative approach to the study of *do*-support. Examining a wider range of possible contexts for *do*-support (i.e. negatives, interrogatives), elucidates observable patterns that previous work has not uncovered. The role played by sentence type is so ubiquitous in the use of *do*-support that cross-tabulation with all other factor groups incorporated into the analysis generated results that helped to address and resolve inconsistencies in the extant literature.

4.5.3 Objective (iii): to demonstrate the value of the *Salem Witchcraft Papers* as a resource for investigating linguistic variation in 17<sup>th</sup> century Massachusetts English (and more broadly, early colonial American English) in its socio-historical context

The strengths of the *Salem Witchcraft Papers* as a resource on 17<sup>th</sup> century English lie in its size and constitution. The relative paucity of *do*-support (given the predominance of affirmative declaratives, where *do*-support is least frequent) means that an exceptionally large amount of data is required to produce enough tokens of *do*-support for effective quantitative analysis. The *Salem Witchcraft Papers* are substantial, providing over 13,000 tokens where *do*-support could have been used or was used, in accordance with the Principle of Accountability (Labov, 1972), a cornerstone of the present study. This also provides access to less frequent constructions (e.g. interrogatives, negatives, and imperatives) which, as demonstrated, play an important role in the development and use of *do*-support. As noted in Sections

2.2 and 3.5.3, the presence of transcription errors in the *Salem Witchcraft Papers* demonstrates a major limitation of this dataset.

Adequate socio-biographical information relating to many of the speakers has survived in the historical record, which offers the researcher the opportunity to correlate social factors with linguistic ones to illuminate sociolinguistic patterns in the data. The data allows insight into the discourse strategies of the predominantly Puritan early colonial society (Archer, 2002; Doty & Hiltunen, 2009; Hiltunen, 1996; Kahlas-Tarkka, 2012; Kahlas-Tarkka & Rissanen, 2011).

The data are also relatively diverse for the time period, including a wide range of text types of differing formality, and giving a voice to those members of early colonial society who are otherwise under-represented in the historical record, including female speakers, those who were lower in the social hierarchy, and those who were enslaved.

The examinations in the data are particularly important in leveraging new insights into the spoken language of the time. Although filtered through a scribal interface, the examinations are closer to actual speech events than other surviving documents, as close to Early Modern English speech as the historical record allows (Kytö & Walker, 2003; Söderlund, 2017: 244-245). This is crucial for the study of linguistic change because change typically originates in speech rather than writing (Rissanen, 1985).

The size of the dataset was crucial for the analysis of lexical collocations, enabling a more detailed and fine-grained investigation than has previously been carried out. 556 different main verbs were identified, including a group of 29 high-frequency verbs (those used more than 100 times). Having a large dataset gave access to a large number of low-frequency verbs which might not have been available in a smaller corpus, and allowed the comparison between verbs of Germanic and Latinate origin (many of which are lower in frequency than Germanic verbs).

The above discussion illustrates many advantages of using the *Salem Witchcraft Papers* as a resource on the use of *do*-support in 17<sup>th</sup> century English, and demonstrates that the corpus is uniquely positioned to



allow detailed investigation of the trajectory of linguistic change involving *do*-support towards the end of the 17<sup>th</sup> century.

## CHAPTER 5: CONTEXTUALIZING QUANTITATIVE PATTERNS

The research presented in the preceding chapters has applied the variationist theoretical framework to the study of two linguistic variables, restrictive relativisation and *do*-support, that were undergoing change at the tail-end of the 17<sup>th</sup> century. Both variables have received a wealth of attention from various areas of linguistics, but there has been a dearth of empirically accountable, quantitative analysis of the variables in Early Modern English speech-like data (but see e.g. Ball, 1996; Bergs, 2005; Díaz-Vera, 1999; Kytö, 1991, 2004, 2022; Nevalainen et al., 2018; Rissanen, 1985, 1997, 2003; Romaine, 1982; Söderlund, 2017; Wischer, 2008). This key consideration provided a major impetus for the research reported in this dissertation.

In contributing to the extant literature, this dissertation targeted several major research objectives, reiterated below:

- I. To reconstruct the variable systems of the target variables (relativisation and *do*-support) in Massachusetts at the end of the 17<sup>th</sup> century, to investigate broader claims in the literature relating to the status of those variable systems in Early American English.
- II. To confirm, refute, and refine (e.g., by targeting methodological infelicities in previous treatments) extant claims in the literature regarding the social and linguistic conditioning of relativisation and *do*-support. Major claims about relativisation include Romaine's (1982) assertion that the *wh*-relatives have made limited inroads into the spoken language, and the claim advanced by Ball (1996) that during the Early Modern period relative markers became subject to an animacy constraint restricting *who* to human antecedents and *which* to non-human antecedents. For *do*-support, particular attention is given to claims regarding the virtual disappearance of affirmative declarative *do* by the end of the 17<sup>th</sup> century (Ellegård, 1953; Kroch, 1989), and lexical effects, where certain verb collocations have been reported to resist the use of *do*-support (Ellegård, 1953; Engblom, 1938; Nurmi, 2018; Warner, 2017).
- III. To evaluate the use of the *Salem Witchcraft Papers* as a linguistic resource for reconstructing antecedent varieties of English, building on the growing literature on the use of historical data for sociolinguistic research (Kytö & Siebers, 2022; López-Couso, 2016; Rissanen, 2003).

Subsequent discussion will summarise the key findings (Section 5.1), and focus on the extent to which the theoretical framework (Section 5.2), dataset (Section 5.3), and variables (Section 5.4) have contributed to meeting the stated research objectives. Section 5.5 identifies several avenues for further research.

## 5.1 Summary of Major Findings

### 5.1.1 Relativisation

Distributional and statistical analysis of relativisation in the *Salem Witchcraft Papers* revealed several patterns. Overall, *that* is the relativiser of choice in 17<sup>th</sup> century English, competing with *which* in subject relatives, and *which* and the zero relative in non-subject relatives. *Which* is the only *wh*-relative to be used to any great extent, in contrast with findings from Present Day English (D’Arcy & Tagliamonte, 2010; Hinrichs et al, 2015; Levey & Hill, 2013), where *who* is the most frequently used *wh*-relative marker. In their study on relativisation in 19<sup>th</sup> century American Civil War Letters, Levey and DeRooy (2021) report much lower rates of *which* than those reported in this dissertation (9% in subject position, 8% in non-subject position). Extrapolating from these findings, it appears that there was a decline in the use of *which* between the 17<sup>th</sup> and 19<sup>th</sup> centuries. This hypothesis is corroborated by the findings in Hinrichs et al. (2015), who observe (in canonically written texts) an increase in relative marker *that* at the expense of *which* between 1960 and 1990. As discussed in more detail in Section 5.5, changes to the relativisation system in the 18<sup>th</sup> and early 19<sup>th</sup> centuries present an intriguing avenue for further investigation.

In subject relatives, *that* predominates in almost all contexts. The only context where another relative comes close is in longer clauses, where *which* is as frequent as *that*. Age plays a notable role in the use of *which*, with younger speakers using *which* significantly more frequently than any other age group. In non-subject relatives, the primary alternation is between *that*, *which*, and the zero relative. In particular, the zero relative is the relativiser of choice in this type of relative clause, while rates for *that* and *which* are broadly similar. Male speakers use zero significantly more than female speakers.

Certain factor groups show consistent effects across subject and non-subject relatives, indicating more general restrictions. This includes the effects of animacy on *wh*-relatives, of processing constraints on zero and the *wh*-relatives, and of proximity to the spoken language.

The results show tentative evidence of an animacy constraint (Ball, 1996). *Who* is too infrequent to permit statistical analysis, but almost categorically co-occurs with human antecedents, while *which* is significantly favoured with non-human antecedents. At this stage, the animacy constraint only really affects the *wh*-relatives. The relative marker *that* is slightly preferred with human antecedents, suggesting that the later association of *that* with non-human antecedents proposed by Ball (1996) is not yet in effect. It is not clear that this shift ever occurs in spoken language: Tagliamonte et al. (2005: 91) report high rates of *that* with both animate and inanimate antecedents in three varieties of Present Day Spoken English (Lowland Scotland, Northern Ireland, and Northwest England).

Processing constraints exert a significant effect over the choice of all relative markers. More cognitively complex constructions are indicated by longer relative clauses, and non-adjacency of the antecedent and the relative marker. Shorter clauses and adjacency of antecedent and relative marker indicate less complex constructions. Less complex constructions favour *that* in subject relatives, and in non-subject favour zero (and to a lesser extent, *that*). The *wh*-relatives, particularly *which*, are favoured in more cognitively complex constructions.

The findings reported here support, to a limited extent, Romaine's (1982) claim that the *wh*-relativisers have not made significant inroads into 17<sup>th</sup> century texts that approximate the spoken language. It should be noted that *who*, *whom*, and *whose* are all very infrequent, even in their putative niche context (i.e. with human antecedents). Document type is used here as a measure of proximity to the spoken language. As records of actual speech events, examinations are the most speech-like of the extant text types. Personal documents (like letters and petitions), and official documents represent the canonically written end of the spectrum, with the latter often featuring specialised, formulaic legal constructions and terminology. Depositions and testimonies occupy the middle of the spectrum. Statistical analysis shows that *which* is significantly favoured in depositions/testimonies and in personal documents, and disfavoured in examinations. Contrastively, *that* and zero are favoured in examinations. This effect is consistent across relative clause types (though the zero relative is mostly absent from subject relatives). *Which* is not absent from more speech-like text types, suggesting that it has begun to infiltrate American speech at the end of the 17<sup>th</sup> century, but is less common in these texts than in canonically written texts.

### 5.1.2 *Do*-Support

As has been well-established in the literature (Ellegård, 1953: 161; Kroch, 1989; Ogura, 1993; Oudesluijs et al., 2022; Söderlund, 2017: 260), the combined contribution of polarity and sentence type is the single most important predictor of the use of *do*-support, and has a bearing on all other factor groups tested. *Do*-support is favoured in interrogatives and negatives, and disfavoured in affirmative declaratives. That being said, there are still cases of affirmative declarative *do*, challenging Kroch's (1989) claim that *do*-support had disappeared from affirmative declaratives by the end of the 17<sup>th</sup> century.

(111) I **doe** verily: beleev: s<sup>d</sup> Ann Puddeater is a: wich

(SWP113.6 - Mary Warren,  
Testimony: 7<sup>th</sup> September, 1692)

Statistical analysis revealed that *do*-support is favoured with weak verbal morphology, and in conjunction with an adverb, and disfavoured in relative clauses. *Do*-support appears to be highly susceptible to social conditioning, as all social factor groups tested were statistically significant. In particular, the use of *do*-support was connected with older speakers, female speakers, and speakers born in America. Consistent with Rissanen's (2003) claims that *do*-support is primarily a feature of the spoken language, *do*-support was favoured in examinations and testimonies, and disfavoured in official and personal documents.

## 5.2 Theoretical Framework

The application of the variationist theoretical framework to the dataset and target variables proved crucial in addressing the key research objectives. Both variables have received notable attention from variationists (for relativisation, see e.g. Ball, 1996; Huber, 2017; Levey & DeRooy, 2021; Romaine, 1982; for *do*-support see e.g. Díaz-Vera, 1999; Rissanen, 1985, 2003; Söderlund, 2017; Wischer, 2008).

In order to reconstruct the variable systems of the target variables, an exhaustive extraction of valid tokens from the data was carried out. Primordial to the variationist approach is the Principle of Accountability (Labov, 1972). Analysing the full range of variants that could have been used allows the researcher to contextualise the findings within the wider variable system, capturing the alternation between variants. This is crucial for establishing the links between the choice of variant and the linguistic context, as well

as the social factors influencing the distribution of variants (Poplack & Tagliamonte, 1989; Wolfram, 1993). In the case of relativisation, this means taking into account more marginal relative markers, including *what*, *as*, *whom*, and *whose*. For *do*-support, the simple form (without *do*-support) is by far the most frequent variant, in part due to the high frequency of affirmative declaratives, where *do*-support is disfavoured. In this instance, failing to include the cases where *do*-support was not used offers a very different (and potentially misleading) picture of variable behaviour.

One of the stated research objectives was to target and verify previous claims in the literature. This included, in certain cases, redressing methodological issues.

For relativisation, Romaine (1982) claims that the *wh*-relatives had made little impact on the spoken language. This claim has drawn criticism, notably from Ball (1996: 243), who states that Romaine's claim was based on data from dialects where *wh*-relatives are infrequent (see also Beal & Corrigan, 2007; Cheshire, 1982). Data from 17<sup>th</sup> century Massachusetts, where the speakers were either British or born in the Americas, serves to put that claim to the test. The *wh*-relatives in the *Salem Witchcraft Papers* are indeed disfavoured in speech-like texts, but still account for 11% of tokens in examinations in both subject and non-subject relatives, suggesting that the *wh*-relatives may have had some impact.

Another example in the analysis of *do*-support is the effect of sentence type, which interacted with every other factor group. Some studies (e.g. Bækken, 2002; Díaz-Vera, 1999; Ihalainen, 1991; Jones & Tagliamonte, 2004; Nurmi, 1999, 2018; Pérez, 1997; Rissanen, 1985; Wischer, 2008; Warner, 2005) limit analysis to only one sentence type, precluding any detailed analysis of the interaction of *do*-support with sentence type. Affirmative declaratives are so frequent in the data that they obscure the effects of other sentence types on *do*-support when the data are examined in the aggregate, and so cross-tabulating sentence type with each other factor group revealed interesting distributional patterns. This was particularly the case with social factor groups: *do*-support appears to not show any differentiation between male and female speakers, but when sex is cross-tabulated with sentence type, it is shown that this is only the case in affirmative declaratives, and there are actually notable differences in the use of affirmative interrogatives (favoured by male speakers) and negative declaratives (favoured by female speakers). Similarly, cross-tabulation of age with sentence type reveals that older speakers predominate in the use of *do*-support in affirmatives, while the youngest speakers are leading in the use of negative declaratives.

One of the major claims about the use of *do*-support is the impact of the *know*-group (Dietze, 1895; Ellegård, 1953; Engblom, 1938; Rohde, 1872), but to my knowledge there has not yet been a detailed quantitative analysis of the effect of specific verb collocations on variant choice, beyond repetitions of the claims made by Ellegård (1953: 201, 206), who, in part, bases these claims on anecdotal observations. Extracting every valid token, in keeping with the Principle of Accountability (Labov, 1972), gives access to the full inventory of main verbs with which *do*-support could have been used in the data. The distributional results supported several divisions along semantic lines: verbs expressing mental processes typically favour *do*-support in interrogatives; verbs expressing speech acts consistently disfavour *do*-support in affirmative declaratives; and emotive verbs favour *do*-support in affirmative declaratives, as do verbs related to legal terminology.

### 5.3 The Salem Witchcraft Papers

Reconstructing historical spoken varieties through the use of diachronic surrogates of orality requires vigorous assessment of the appropriateness of the data. The choice of dataset is a vital component of the research process as this places important constraints on what is discoverable. Working with historical varieties severely limits the researcher to the data that has survived, which may be fragmented, limited in quantity (hampering the use of quantitative methodologies), or skewed in favour of certain groups in society, typically educated, higher- status males (Romaine, 2016).

The *Salem Witchcraft Papers* are an indispensable resource for the study of 17<sup>th</sup> century English, affording detailed elucidation of Early American English. The people represented in the trial data are relatively diverse, and the examination records in particular give a voice to many of the (predominantly female) people accused of witchcraft. The importance of this cannot be understated, and is highlighted by the statistically significant influence of social factor groups on the use of relatives (sex and document type) and *do*-support (where sex, age, birthplace, and document type were all statistically significant predictors of the use of *do*-support). The data also features a wide age-range of speakers, from young children to people in their 70s and older. A major advantage of this is it allows for the tracking of apparent-time changes, under the assumption that a change in progress will be reflected in the language of different age groups (Bailey et al., 1992; Chambers & Trudgill, 1980: 165; Labov, 1963). This effect can be seen in the

results for both variables: younger speakers use *wh*-relatives more frequently than older speakers; and older speakers use more affirmative declarative *do* than younger speakers.

It is important to understand that any written record, no matter how accurate, is fundamentally distinct from the original speech act. Filtered through a scribal interface, the speech-like data in the *Salem Witchcraft Papers* is subject to editorial interference. Nonetheless, a working hypothesis is that written texts attempting to record spoken dialogue are the best source of available data to reconstruct diachronic orality, when interpreted judiciously (Auer et al., 2015: 7; Kytö & Walker, 2003: 221).

#### 5.4 The Target Variables

The linguistic variables targeted in this dissertation were not chosen at random. Both variables have been reported to be undergoing change during the end of the 17<sup>th</sup> century (Ball, 1996; Ellegård, 1953; Johansson, 2017; Rissanen, 1985; Söderlund, 2017). Both variables have received notable attention in the literature, though, as this dissertation has been at pains to point out, not always in ways that are methodologically optimal or unimpeachable. Both variables have putative origins in Old English, though it is the transition from Middle English to Early Modern English that is responsible for the changes reported to be in progress during the 17<sup>th</sup> century, such as the rise of the *wh*-relatives (Romaine, 1982) and the establishment of an animacy constraint (Ball, 1996), the decline of subject zero (Johansson, 2017; Tottie & Johansson, 2015), the rise of *do*-support in negatives and interrogatives, and the decline in affirmative declarative *do* (Ellegård, 1953).

The (often statistically significant) social embedding of relativisation and *do*-support demonstrates the importance of situating diachronic research in its broader socio-historical context. Text type, operationalised here as a measure of formality and proximity to speech, played a significant role in the selection of variants for both variables. Examinations consistently favoured the variants associated most closely with the spoken language (e.g. the zero relative, *do*-support). Canonical written documents favoured the use of the *wh*-relatives. Interestingly, *do*-support was most frequent in examinations and official documents, the most and least speech-like text types respectively. However, this is attributed to the prevalence of affirmative declarative *do* in official documents. The working hypothesis is that, as the written language is typically more conservative than the spoken, less speech-like text types have retained



the use of *do*-support in affirmative declaratives, where it has otherwise declined in more speech-like text types.

### 5.5 Future Avenues for Investigation

To conclude, I will tentatively suggest some avenues for further investigation. To gain a more comprehensive picture of the development of each of the target variables, subsequent work could focus on the periods directly preceding and following the *Salem Witch Trials*. A lack of surviving data may hamper attempts to look at relativisation and *do*-support in the early to mid-17<sup>th</sup> century. One possible option for Early American English is the texts examined by Rissanen (1985), which cover the period from 1630-1650 and include records of church meetings and historical chronicles. The 18<sup>th</sup> century has curiously attracted less attention in the literature (Huber, 2017), but will likely offer interesting insight into the development of these two variables. As mentioned previously, 19<sup>th</sup> century data (Levey & DeRooy, 2021) is suggestive of a decline in the use of *which* in American English at some point during the 18<sup>th</sup> century. An approach that is methodologically consistent with the current dissertation would allow for direct comparison of findings.

Another approach is to utilise the *Salem Witchcraft Papers* for other variables that are undergoing change at the end of the 17<sup>th</sup> century. Rissanen (2003) presents some groundwork for this line of inquiry, conducting preliminary analyses of the alternation between the auxiliaries *shall* and *will*, and variation in negation.

A further possibility is to tap into the insights offered by 18<sup>th</sup> century grammarians, who highlight a number of linguistic features that they consider ‘improper’. As a relevant example, the use of *that* as a relative marker is criticised, particularly with human antecedents, as is the use of *which* with human antecedents, and *who* with non-human antecedents (Leonard, 1962: 261-262). The zero relative is seen as improper, even in non-subject relatives (Leonard, 1962: 281-282).

The *Salem Witchcraft Papers* reveals extensive evidence of variable verbal morphology, particularly in the past tense. This is evidenced in the chapter on *do*-support, where there are cases of verbs that have strong verbal morphology in standard Present Day English, but appear with weak verbal morphology in

the data (e.g. *caught* vs. *caught*, *telled* vs. *told*, *seed* vs *saw*) (Cheshire, 1994: 116; Lass, 1994). This could present a fruitful topic for further investigation, particularly as such variability also drew the attention of the 18<sup>th</sup> century grammarians (Leonard, 1962: 268-271).

Notwithstanding the direction taken in future investigations, the research presented in this dissertation demonstrates the value in comprehensive, accountable analysis of linguistic variation. The variation between semantically equivalent forms is systematic, not random, and the variationist framework is uniquely positioned to elucidate underlying structured patterns of variation. The only viable access to the structured heterogeneity of historical spoken varieties is through the judicious use of appropriate textual evidence to reconstruct the speech of the past in its socio-historical context. It is in the pursuit of diachronic orality that the indispensable value of the *Salem Witchcraft Papers* as a linguistic resource clearly comes to the fore.

## BIBLIOGRAPHY

- Adamson, H. D. (1992). Social and processing constraints on relative clauses. *American Speech*, 67(2), 123-133.
- Akahane, H. (1998). Do-support and tier scansion: A minimalist account of HMC effects. *English Linguistics*, 15, 81-110.
- Alcázar, A., & Saltarelli, M. (2013). *The syntax of imperatives*. Cambridge: Cambridge University Press.
- Al-Maktrah, E. S., Yasin, M. S. M., Sulaiman, M. Z., & Al-Khawaldeh, M. (2017). The markedness in acquiring do-support in negation and inversion by adult native-Arabic speakers learning English. *Journal of Language Teaching, Linguistics, and Literature*, 23(4), 143-156.
- Anderssen, M., & Bentzen, K. (2018). Different outcomes in the acquisition of residual V2 and do-support in three Norwegian-English bilinguals: Cross-linguistic influence, dominance and structural ambiguity. *Frontiers in Psychology*, 9, 1-18.
- Archer, D. (2002). "Can innocent people be guilty?": A sociopragmatic analysis of examination transcripts from the Salem witchcraft trials. *Journal of Historical Pragmatics*, 3(1), 1-30.
- Archer, D. (2005). *Questions and answers in the English courtroom (1640-1760): A sociopragmatic analysis*. Amsterdam: John Benjamins Publishing.
- Ard, J. (1982). Auxiliary *do*: Support or emphasis? *Linguistics*, 20, 445-466.
- Ariel, M. (1990). *Accessing noun phrase antecedents*. London: Routledge.
- Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, W. Spooren (eds.) *Text representation: Linguistic and psycholinguistic aspect* (29-88). Amsterdam: John Benjamins.
- Arnold, D. (2007). Non-restrictive relatives are not orphans. *Journal of Linguistics*, 43(2), 271-309.
- Auer, A., Peersman, C., Pickl, S., Rutten, G., & Vosters, R. (2015). Historical sociolinguistics: the field and its future. *Journal of Historical Sociolinguistics*, 1(1), 1-12.
- Ayres-Bennett, W. (2004). *Sociolinguistic variation in seventeenth-century France: Methodology and case studies*. Cambridge: Cambridge University Press.
- Bækken, B. (2002). Yet this follie doth many times assault the brauest minds: Affirmative declarative *do* in 17<sup>th</sup>-century English. *Nordic Journal of English Studies*, 1(2), 317-337.
- Bahns, J. (1991). What did you bought? Explaining a typical error in the acquisition of English. *International Review of Applied Linguistics in Language Teaching*, 29(3), 213-228.
- Bailey, C. J. N. (1973). *Variation and linguistic theory*. Arlington: Center for Applied Linguistics.
- Bailey, G., Wikle, T., Tillery, J., & Sand, L. (1992). The apparent time construct. *Language Variation and Change*, 3, 241-264.
- Baker, E. W. (2015). *A storm of witchcraft: The Salem trials and the American experience*. Oxford: Oxford University Press.
- Ball, C. N. (1994). The origins of the informative-presupposition it-cleft. *Journal of Pragmatics*, 22(6), 603-628.

- Ball, C. N. (1996). A diachronic study of relative markers in spoken and written English. *Language variation and change*, 8(2), 227-258.
- Bayley, R. (1999). Relativization strategies in Mexican-American English. *American Speech*, 74(2), 115-139.
- Beal, J. C., & Corrigan, K. P. (2007). 'Time and Tyne': A corpus-based study of variation and change in relativization strategies in Tyneside English. In S. Elspaß, N. Langer, J. Scarloth, & W. Vandebussche (eds.) *Germanic language histories 'from below' (1700-2000)* (99-114). Berlin: De Gruyter Mouton.
- Beaman, K. (1984). Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse. In D. Tannen (ed.), *Coherence in spoken and written discourse* (45-80). New Jersey: Ablex.
- Bengtsson, P. (1996). Prepositional relatives in English: A diachronic case study of three authors. [Unpublished paper, Department of English, Göteborg University].
- Benincà, P., & Poletto, C. (2004). A case of do-support in Romance. *Natural Language and Linguistic Theory*, 22(1), 51-94.
- Bergh, G., & Seppänen, A. (2000). Preposition stranding with wh-relatives: A historical survey. *English Language & Linguistics*, 4(2), 295-316.
- Bergs, A. (2005). *Social networks and historical sociolinguistics: Studies in morphosyntactic variation in the Paston letters (1421-1503)*. Berlin: De Gruyter Mouton.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Blumberg, J. (2022, October 24). A brief history of the Salem witch trials. *Smithsonian.com*.  
<https://www.smithsonianmag.com/history/a-brief-history-of-the-salem-witch-trials-175162489/>.
- Boyer, P., & Nissenbaum, S. (1977). *Salem Witchcraft Papers: Verbatim transcripts of the legal documents of the Salem witchcraft outbreak*. New York: Da Capo Press.
- Braddick, M. J. (2009). Introduction: The politics of gesture. *Past & Present*, 4, 9-35.
- Britain, D. (2020). What happened to those relatives from East Anglia? A multilocality analysis of dialect levelling in the relative marker system. In K.V. Beaman, I. Buchstaller, S. Fox, & J.A. Walker (eds.) *Advancing sociogrammatical variation and change* (93-114). London: Routledge.
- Brock-Nannestad, G., & Fontaine, J. M. (2008). Early use of the Scott-Koenig phonautograph for documenting performance. *Journal of the Acoustical Society of America*, 123(5), 6239-6244.
- Bruening, B. (2010). Language-particular syntactic rules and constraints: English locative inversion and do-support. *Language*, 86(1), 43-84.
- Burns, M., & Rosenthal, B. (2008). Examination of the records of the Salem witch trials. *The William and Mary Quarterly*, 65(3), 401-422.
- Bybee, J. L., & Beckner, C. (2010). Usage-based theory. In B. Heine, & H. Narrog (eds.), *The Oxford handbook of linguistic analysis* (827-855). Oxford: Oxford University Press.
- Chambers, J. K., & Trudgill, P. (1980). *Dialectology*. Cambridge: Cambridge University Press.
- Cheshire, J. (1982). *Variation in an English dialect: A sociolinguistic study*. Cambridge: Cambridge

University Press.

- Cheshire, J. (1994). Standardization and the English irregular verbs. In D. Stein, I. Tieken-Boon Van Ostade (eds.) *Towards a standard English: 1600-1800* (115-133). Berlin: De Gruyter Mouton.
- Cheshire, J. (2005). Syntactic variation and beyond: Gender and social class variation in the use of discourse-new markers. *Journal of Sociolinguistics*, 9(4), 479-508.
- Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, 2(3), 113-124.
- Craker, W. D. (1997). Spectral evidence, non-spectral acts of witchcraft, and confession at Salem in 1692. *The Historical Journal*, 40(2), 331-358.
- Cuesta, J. F. (2016). The loss of the Northern subject rule and the rise of the DO-periphrasis in 16<sup>th</sup> century Northern English wills: A possible case of exaptation. *Neuphilologische Mitteilungen*, 117(1), 109-136.
- Culicover, P. W. (2008). The rise and fall of constructions and the history of English *do*-support. *Journal of Germanic Linguistics*, 20(1), 1-52.
- Culpeper, J., & Archer, D. (2008). Requests and directness in Early Modern English trial proceedings and play texts, 1640-1760. *Pragmatics & Beyond*, 176, 45-84.
- Culpeper, J., & Kytö, M. (2000). Data in historical pragmatics: Spoken interaction (re) cast as writing. *Journal of Historical Pragmatics*, 1(2), 175-199.
- Culpeper, J., & Kytö, M. (2010). *Early modern English dialogues: Spoken interaction as writing*. Cambridge: Cambridge University Press.
- D'Arcy, A., & Tagliamonte, S. A. (2010). Prestige, accommodation, and the legacy of relative *who*. *Language in Society*, 39(3), 383-410.
- De Both, F. (2019). Nonstandard periphrastic DO and verbal -s in the south west of England. *Journal of Historical Sociolinguistics*, 5 (1), 1-35.
- Dekeyser, X. (1984). Relativizers in Early Modern English. In J. Fisiak (ed.) *Historical syntax*, (61-88). Berlin: De Gruyter Mouton.
- Dekeyser, X., & Ingels, M. (1988). Socio-historical aspects of relativisation in late 16<sup>th</sup> century English: ca. 1550-1600. *Studia Anglica Posnaniensia*, 21, 25-39.
- Demos, J. (1983). *Entertaining Satan: Witchcraft and the culture of early New England*. Oxford: Oxford University Press.
- Denison, D. (1993). *English historical syntax*. London: Routledge.
- Deuber, D. (2014). *English in the Caribbean: Variation, style and standards in Jamaica and Trinidad*. Cambridge: Cambridge University Press.
- De Villiers, J. G., Tager Flusberg, H. B., Hakuta, K., & Cohen, M. (1979). Children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, 8(5), 499-518.
- De Vries, M. (2002). *The syntax of relativization*. Utrecht: LOT.
- Díaz-Vera, J. E. (1999). Reconstructing variation in the speech community: Evidence on early American English negative constructions from the *Salem Witchcraft Papers*. *Cuadernos de Filología*

*Inglesa*, 8, 275-289.

- Diessel, H., & Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language*, 81(4), 882-906.
- Dietze, H. (1895). *Das umschreibende 'do' in der neuenglischen Prosa*. Germany: Frommansche Hof-Buchdr.
- Doty, K. L. (2007). Telling tales: The role of scribes in constructing the discourse of the Salem witchcraft trials. *Journal of Historical Pragmatics*, 8(1), 25-41.
- Doty, K. L., & Hiltunen, R. (2002). "I will tell, I will tell": Confessional patterns in the Salem witchcraft trials, 1692. *Journal of Historical Pragmatics* 3(2), 299-335.
- Doty, K. L., & Hiltunen, R. (2009). Formulaic discourse and speech acts in the witchcraft trial records of Salem, 1692. *Journal of Pragmatics*, 41(3), 458-469.
- Eitelmann, M. (2016). Support for end-weight as a determinant of linguistic variation and change. *English Language and Linguistics*, 20(3), 395-420.
- Ellegård, A. (1953). *The auxiliary 'do': The establishment and regulation of its use in English*. Stockholm, Sweden: Almqvist & Wiksell.
- Engblom, V. (1938). *Origin and early development of the auxiliary do*. Lund: Berlingska Brktryckeriet.
- Fischer, O. (1992). Syntax. In N. Blake (ed.), *The Cambridge history of the English language, vol. II: 1066-1476*, (207-408). Cambridge: Cambridge University Press.
- Fischer, O. (2008). Syntax. In W. van der Wurff, D. Denison, & R. Hogg (eds.), *A history of the English language*, (109-198). Cambridge: Cambridge University Press.
- Fischer, O., Van Kemenade, A., Koopman, W., & Van der Wurff, W. (2000). *The syntax of early English*. Cambridge: Cambridge University Press.
- Fitzmaurice, S. M., & Taavitsainen, I. (2007). *Methods in historical pragmatics*. Berlin: Mouton de Gruyter.
- Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The cumulative-enhancement model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism*, 1(1), 3-16.
- Fowler, H. W., & Fowler, F. G. (1931). *The concise Oxford dictionary of current English*. Oxford: The Clarendon Press.
- Geisler, C. (2002). Relativization in Ulster English. In P. Poussa (ed.) *Relativization on the North Sea littoral* (135-146) Munich: LINCOM Europa.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1-76.
- Görlach, M. (1991). *Introduction to early modern English*. Cambridge: Cambridge University Press.
- Gotthard, L. (2019). Why *do*-support in Scots is different. *English Studies*, 100(3), 314-338.
- Grund, P. (2007). From tongue to text: The transmission of the Salem witchcraft examination records. *American Speech*, 82(2), 119-150.
- Grund, P. J. (2012). Textual history as language history? Text categories, corpora, editions, and the witness depositions from the Salem witch trials. *Studia Neophilologica*, 84(1), 40-54.

- Grund, P. (2021). *The sociopragmatics of stance: Community, language, and the witness depositions from the Salem witch trials*. Amsterdam: John Benjamins.
- Grund, P., Kyto, M., & Rissanen, M. (2004). Editing the Salem witchcraft records: An exploration of a linguistic treasury. *American Speech*, 79(2), 146-167.
- Grund, P. J., Burns, M., & Peikola, M. (2014). The vagaries of manuscripts from the Salem witch trials: An edition of four (re-) discovered documents from the case against Margaret Scott of Rowley. *Studia Neophilologica*, 86(1), 37-50.
- Guasti, M. T., & Cardinaletti, A. (2003). Relative clause formation in Romance child production. *Probus*, 15(1), 47-89.
- Guy, G. R. (1989). Language and social class. In F. J. Newmeyer (ed.), *Linguistics: The Cambridge survey*, (37-63). Cambridge: Cambridge University Press.
- Guy, G. R., & Bayley, R. (1995). On the choice of relative pronouns in English. *American Speech*, 70(2), 148-162.
- Guy, G., Horvath, B., Vonwiller, J., Daisely, E., & Rogers, I. (1986). An intonational change in progress in Australian English. *Language in Society*, 15(1), 23-51.
- Hackenberg, R. (1972). *Appalachian English: A sociolinguistic study*. [Unpublished Ph.D. thesis, Georgetown University].
- Haddican, B. (2007). On Egin: Do-support and VP focus in Central and Western Basque. *Natural Language and Linguistic Theory*, 25(4), 735-764.
- Han, C., & Kroch, A. (2000). The rise of *do*-support in English: Implication for clause structure. *North East Linguistics Society*, 30(23), 1-15.
- Harvie, D. M. (1998). *Tracking down old relatives: Zero relatives in subject and non-subject function in Early African American English*. [Master's thesis, University of Ottawa]. University of Ottawa Library.
- Haskins, G. L. (1954). Codification of the law in Colonial Massachusetts: A study in comparative law. *Indiana Law Journal*, 30(1), 1-17.
- Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language*, 75(2), 244-285.
- Hazen, K. (2007). The study of variation in historical perspective. In R. Bayley, & C. Lucas, *Sociolinguistic variation*, (70-89). Cambridge: Cambridge University Press.
- Hein, J. (2017). Doubling and *do*-support in verbal fronting: Towards a typology of repair operations. *Glossa: A Journal of General Linguistics*, 2(1), 1-36.
- Herrmann, T. (2005). Relative clauses in English dialects of the British Isles. In B. Kortmann, T. Herrmann, L. Pietsch, & S. Wagner (eds.) *A comparative grammar of British English dialects* (21-124). Berlin: De Gruyter Mouton.
- Hewson, J. (1990). The auxiliary *DO* in English. *Linguistica Atlantica*, 12, 39-52.
- Hill, C. (2018). *Society and puritanism in pre-Revolutionary England*. London: Verso Books.
- Hiltunen, R. (1996). "Tell me, be you a witch?": Questions in the Salem witchcraft trials of 1692.

- International Journal for the Semiotics of Law*, 9(25), 17-37.
- Hiltunen, R. (2010). Discourse and context in a historical perspective: On courtroom interaction in Salem, 1692. In M. Johansson, M. Raitaniemi, M. L. Helasvuo, & S. K. Tanskanen (eds.), *Discourses in interaction*, (61-78). Amsterdam: John Benjamins.
- Hiltunen, R., & Peikola, M. (2007). Trial discourse and manuscript context: Scribal profiles in the Salem witchcraft records. *Journal of Historical Pragmatics*, 8(1), 43-68.
- Hinrichs, L., Szmrecsanyi, B., & Bohmann, A. (2015). "Which"-hunting and the standard English relative clause. *Language*, 91(4), 806-836.
- Hoekstra, J. (2016). Beyond do-support and tun-periphrasis: The case of finite verb doubling in Karrharde North Frisian. *Journal of Germanic Linguistics*, 28, 317-348.
- Hoffer, P. C. (1997). *The Salem witchcraft trials: A legal history*. Lawrence, Kansas: University Press of Kansas.
- Huber, M. (2017). Structural and sociolinguistic factors conditioning the choice of relativizers in Late Modern English: A diachronic study based on the Old Bailey Corpus. *Nordic Journal of English Studies*, 16(1), 74-119.
- Huddleston, R. D. (1971). *The sentence in written English: A syntactic study based on an analysis of scientific texts*. Cambridge: Cambridge University Press.
- Huddleston, R. D. (1976). *Introduction to the grammar of English*. Cambridge: Cambridge University Press.
- Hundt, M. (2009). Colonial lag, colonial innovation or simply language change. In G. Rohdenburg, & J. Schlüter (eds.), *One language, two grammars? Differences between British and American English*, (13-37). Cambridge: Cambridge University Press.
- Hundt, M. (2015). Do-support in early New Zealand and Australian English. In P. Collins (ed.), *Grammatical change in English world-wide*, 65-86. Amsterdam: John Benjamins.
- Ihalainen, O. (1976). Periphrastic 'do' in affirmative sentences in the dialect of East Somerset. *Neuphilologische Mitteilungen* 77(4), 608-622.
- Ihalainen, O. (1980). Relative clauses in the dialect of Somerset. *Neuphilologische Mitteilungen*, 81(2), 187-196.
- Ihalainen, O. (1991). Periphrastic do in affirmative sentences in the dialect of East Somerset. In P. Trudgill, J. K. Chambers (eds.), *Dialects of English: Studies in grammatical variation*, 104-119. Abingdon: Routledge.
- Ingels, M. B. (1985). *Socio-historical aspects of relativization in 16<sup>th</sup> century English (ca.1550-1600)*. [Unpublished PhD dissertation, K. U. Leuven].
- Jaeger, T. F., & Wasow, T. (2005). Processing as a source of accessibility effects on variation. In R. T. Cover (ed.) *Proceedings of the 31st annual meeting of the Berkeley Linguistics Society* (169-180). Berkeley: Berkeley Linguistics Society.
- Johansson, C. (2006) Relativizers in nineteenth-century English. In M. Kyto, M. Rydén, & E. Smitterberg (eds.), *Nineteenth-century English: Stability and change*, (136-182). Cambridge: Cambridge



University Press.

- Johansson, C. (2012). Relativization in Early Modern English: Written versus speech-related genres. In L. Brinton & A. Bergs (eds) *English historical linguistics: An international handbook*, (776–790). Berlin: De Gruyter Mouton.
- Johansson, C. (2017). Relativization. In A. Bergs & L. J. Brinton (eds.) *The History of English*, (267-286). Berlin: De Gruyter Mouton.
- Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, 3(1), 359-383.
- Jones, M., & Tagliamonte, S. (2004). From Somerset to Samaná: Preverbal did in the voyage of English. *Language Variation and Change*, 16(2), 93-126.
- Jucker, A. H. (1995). *Historical pragmatics. Pragmatic developments in the history of English*. Amsterdam: John Benjamins.
- Jucker, A. H., & Taavitsainen, I. (2013). *English historical pragmatics*. Edinburgh: Edinburgh University Press.
- Kahlas-Tarkka, L. (2012). “I am a Gosple Woman”: On language in the courtroom discourse during the Salem witch trials, with special reference to female examinees. *Studia Neophilologica*, 84(1), 55-69.
- Kahlas-Tarkka, L., & Rissanen, M. (2011). Attitudes of the accused in the Salem witchcraft trials. In P. Pahta, & A. H. Jucker, *Communicating early English manuscripts*, (241-258). Cambridge: Cambridge University Press.
- Kallel, A. (2002). The age variable in the rise of periphrastic ‘do’ in English. *Reading Working Papers in Linguistics*, 6, 161-185.
- Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8(1), 63-99.
- Kerswill, P. (2009). Language and social class. In J. Culpeper, F. Katamba, P. Kerswill, A. McEnery, R. Wodak & T. McEnery, (eds.) *English language* (358-372). Basingstoke: Palgrave Macmillan.
- Klemola, J. (2018). The historical geographical distribution of periphrastic DO in southern dialects. In L. White (ed.), *Southern English varieties then and now*, (262-292). Berlin: De Gruyter Mouton.
- Kocić, A. (2010). Salem witchcraft trials: The perception of women in history, literature and culture. *Facta Universitatis* 8(1), 1-7.
- Kortmann, B. (2004). *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*. Berlin: De Gruyter Mouton.
- Kroch, A. (1989). Function and grammar in the history of English: Periphrastic *do*. In R. W. Fasold & D. Schiffirin (eds.), *Language change and variation*, 134-171. Amsterdam: John Benjamins.
- Kryk-Kastovsky, B. (2000). Representations of orality in Early Modern English trial records. *Journal of Historical Pragmatics*, 1(2), 201-230.
- Kurath, H. (1986). The sociocultural background of dialect areas in American English. In H. B. Allen, &

- M. D. Linn (eds.), *Dialect and language variation*, (98-116). Amsterdam: Elsevier.
- Kytö, M. (1991). Can (could) vs. may (might): Regional variation in Early Modern English. In D. Kastovsky (ed.), *Historical English syntax*, (233-289). Berlin: De Gruyter Mouton.
- Kytö, M. (2004). The emergence of early American English: Evidence from seventeenth-century records in New England. In R. Hickey (ed.), *Legacies of colonial English: Studies in transported dialects*, (121-157). Cambridge: Cambridge University Press.
- Kytö, M. (2022). Coordination in the courtroom: The uses of AND in the records of the Salem witchcraft trials. In M. Kytö, & L. Siebers (eds.), *Earlier North American Englishes*, (37-63). Amsterdam: John Benjamins.
- Kytö, M., & Siebers, L. (2022). *Earlier North American Englishes*. Amsterdam: John Benjamins.
- Kytö, M., & Walker, T. (2003). The linguistic study of Early Modern English speech-related texts: How “bad” can “bad” data be? *Journal of English Linguistics*, 31(3), 221-248.
- Kytö, M., Walker, T., & Grund, P. (2007). English witness depositions 1560-1760: An electronic text edition. *ICAME Journal* (31), 65-85.
- Labov, W. (1963). The social motivation of a sound change. *Word*, 19(3), 273-309.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, W. (1969). *The study of nonstandard English*. Washington, DC: National Council of Teachers of English.
- Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, 1(1), 97-120.
- Labov, W. (1974). Linguistic change as a form of communication. In M. Silverstein (ed.) *Human communication: theoretical explanations*, (221-256). London: Routledge.
- Labov, W. (1980). *Locating language in time and space*. New York: Academic Press.
- Labov, W. (1981). What can be learned about change in progress from synchronic description? In D. Sankoff & H. Cedergren (eds.) *Variation omnibus*, (177-201). Edmonton: Linguistic Research.
- Labov, W. (1982). Building on empirical foundations. In W. Lehmann, & Y. Malkiel (eds.), *Perspectives on historical linguistics*, (17-92). Amsterdam: John Benjamins.
- Labov, W. (1990). The interaction of sex and social class in the course of linguistic change. *Language Variation and Change*, 2(2), 205-254.
- Labov, W. (1994). *Principles of linguistic change. Volume 1: Internal factors*. Oxford: Blackwell.
- Langer, N. (2001). *Linguistic purism in action: How auxiliary tun was stigmatized in early New High German*. Berlin: De Gruyter Mouton.
- Laskaris, I. (2019). Agency and emotion of young female accusers in the Salem witchcraft trials. *Cultural and Social History*, 16(4), 413-429.
- Leonard, S. A. (1962). *The doctrine of correctness in English usage. 1700-1800*. New York: Russell & Russell.
- Levey, S. (2017). Three centuries of restrictive relativization: Evidence from The Corpus of English Dialogues and the Old Bailey Proceedings (1600-1900). Paper presented at SHEL 10, University of Kansas.

- Levey, S., & DeRooy, G. (2021). North versus South: Regional patterns of grammatical variation in nineteenth-century American English. *American Speech*, 96(3), 297-331.
- Levey, S., & Hill, C. (2013). Social and linguistic constraints on relativizer omission in Canadian English. *American speech*, 88(1), 32-62.
- Lindelöf, A. (1997). Prepositional relatives in English: A diachronic study of *which* and *who(m)* in Early Modern English texts. [Unpublished paper. Department of English, Göteborg University].
- López-Couso, M. J. (2016). Continuing the dialogue between corpus linguistics and grammaticalization theory: Three case studies. *Corpus Linguistics and Linguistic Theory*, 12(1), 7-29.
- Los, B. (2015). *A historical syntax of English*. Edinburgh: Edinburgh University Press.
- Macaulay, R. K. S. (1991). *Locating dialect in discourse: The language of honest men and bonnie lasses in Ayr*. Oxford: Oxford University Press.
- Mallinson, C. (2007). Social class, social status and stratification: Revisiting familiar concepts in sociolinguistics. *University of Pennsylvania Working Papers in Linguistics*, 13(2), 149-163.
- Marckwardt, A. H. (1958). *American English*. Oxford: Oxford University Press.
- McWhorter, J. H. (2009). What else happened to English? A brief for the Celtic hypothesis. *English Language and Linguistics*, 13(2), 163-191.
- Milroy, J. (1992). *Linguistic variation and change: On the historical sociolinguistics of English*. Oxford: Blackwell.
- Moessner, L. (2017). Standardization. In A. Bergs, & L. J. Brinton (eds.), *The history of English: Early modern English*, 167-187. Berlin: De Gruyter Mouton.
- Molencki, R. (1987). Some observations on relative clauses in the Old English version of Bede's "Historia Ecclesiastica Gentis Anglorum". *Studia Anglica Posnanensia*, 20, 83-99.
- Montgomery, M. (1997). Making transatlantic connections between varieties of English: The case of plural verbal-s. *Journal of English Linguistics*, 25(2), 122-141.
- Mustanoja, T. F. (1960). *A middle English syntax*. Helsinki: Société Néophilologique.
- Nevalainen, T. (1999). Making the best use of 'bad' data: evidence for sociolinguistic variation in Early Modern English. *Neuphilologische Mitteilungen*, 100(4), 499-533.
- Nevalainen, T. (2006). *Introduction to early modern English*. Edinburgh: Edinburgh University Press.
- Nevalainen, T., Palander-Collin, M., & Säily, T. (2018). *Patterns of change in 18<sup>th</sup>-century English*. Amsterdam: John Benjamins.
- Nevalainen, T., & Raumolin-Brunberg, H. (2017). *Historical sociolinguistics: Language change in Tudor and Stuart England*. Abingdon: Routledge.
- Nevalainen, T., & Rissanen, M. (1986). Do you support the *do*-support? Emphatic and non-emphatic *DO* in affirmative statements in present-day spoken English. In S. Jacobson (ed.), *Papers from the third Scandinavian symposium on syntactic variation*, (35-50). Stockholm: Almqvist & Wiksell.
- Nurmi, A. (1999). Auxiliary *do* in fifteenth-century English: Dialectal variation and formulaic use. In I. Taavitsainen, G. Melchers, & P. Pahta (eds.), *Writing in nonstandard English*, (225-242). Amsterdam: John Benjamins.

- Nurmi, A. (2000). The rise and fall of periphrastic *DO* in early modern English, or “Howe the Scotts will declare themself’s”. In R. Bermudez-Otero, D. Denison, R. M. Hogg, & C. B. McCully (eds.), *Generative theory and corpus studies: A dialogue from 10 ICEHL*, (373-394). Berlin: De Gruyter Mouton.
- Nurmi, A. (2018). Periphrastic *do* in eighteenth-century correspondence. In T. Nevalainen, M. Palander-Collin, & T. Säily (eds.), *Patterns of change in 18<sup>th</sup>-century English. A sociolinguistic approach*, (117-135). Amsterdam: John Benjamins.
- Ogura, M. (1993). The development of periphrastic *do* in English: A case of lexical diffusion in syntax. *Diachronica*, 10(1), 51-85.
- Oudesluijs, T., Gordon, M., & Auer, A. (2022). Urbanisation, supralocalisation and the development of periphrastic *DO* in early modern England. *Journal of Historical Syntax*, 6(16), 1-25.
- Owen, J. (2018). An evaluation of the Celtic hypo’ for Brythonic Celtic influence on early English. (Master’s Dissertation, The Australian National University).
- Pargman, S. (2004). Gullah *duh* and periphrastic *do* in English dialects: Another look at the evidence. *American Speech*, 79(1), 3-32.
- Partridge, A. C. (1948). The periphrastic auxiliary verb “*do*” and its use in the plays of Ben Jonson. *The Modern Language Review*, 43(1), 26-33.
- Peikola, M. (2012). Supplicatory voices: Genre properties of the 1692 petitions in the Salem witch trials. *Studia Neophilologica*, 84(1), 106-118.
- Pérez, J. R. V. (1997). The use of periphrastic “*do*” in early modern English negative declaratives: Evidence from the Helsinki Corpus. *SEDERI*, 8, 35-43.
- Pfenninger, S. E. (2014). Quadrilingual advantages: Do-support in bilingual vs. multilingual learners. *International Journal of Multilingualism*, 11(2), 143-163.
- Piercy, C., Rickford, J., Wasow, T., Jaeger, T. F., and Thomas, E. (2011). Language processing and relative clauses in Dorset English. Paper presented at Variation and Language Processing 2011, Chester, U.K., Apr. 11–13.
- Pintzuk, S. (2017). Variationist approaches to syntactic change. In B. D. Joseph, & R. D. Janda (eds.), *The handbook of historical linguistics*, (509-528). Oxford: Blackwell.
- Poplack, S. (2018). *Borrowing: Loanwords in the speech community and in the grammar*. Oxford: Oxford University Press.
- Poplack, S., & Levey, S. (2010). Contact-induced grammatical change: A cautionary tale. In P. Auer & J. E. Schmidt (eds.) *Language and space – an international handbook of linguistic variation. Vol 1: Theories and methods*, (391-419). Berlin, Germany: De Gruyter Mouton.
- Poplack, S., & St-Amand, A. (2007). A real-time window on 19<sup>th</sup>-century vernacular French: The Récits du français québécois d'autrefois. *Language in Society*, 36(5), 707-734.
- Poplack, S., & Tagliamonte, S. (1989). There’s no tense like the present: Verbal -s inflection in early Black English. *Language Variation and Change*, 1(1), 47-84.
- Poplack, S., & Tagliamonte, S. (2001). *African American English in the diaspora*. Oxford: Blackwell.

- Poplack, S., Walker, J. A., & Malcolmson, R. (2006). An English “like no other”? Language contact and change in Quebec. *Canadian Journal of Linguistics*, 51(2-3), 185-213.
- Poplack, S., Zentz, L., & Dion, N. (2012). Phrase-final prepositions in Quebec French: An empirical study of contact, code-switching and resistance to convergence. *Bilingualism: Language and Cognition*, 15: 203-225.
- Poussa, P. (1990). A contact-universals origin for periphrastic do, with special consideration of OE Celtic contact. In *Papers from the fifth international conference on English historical linguistics*, (407-434). Amsterdam: John Benjamins.
- Preusler, W. (1956). Keltischer Einfluss im Englischen. *Revue des Langues Vivantes*, 22, 322-350.
- Quirk, R. (1957). Relative clauses in educated spoken English. *English Studies*, 38, 97-109.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Ray, B. (2003). Salem witch trials. *OAH Magazine of History*, 17(4), 32-36.
- Ray, B. (2018). *Salem witchcraft papers*. Salem Witch Trials Documentary Archive and Transcription Project. <https://salem.lib.virginia.edu/category/swp.html>
- Rickford, J. (1986). Social contact and linguistic diffusion: Hiberno-English and new world Black English. *Language*, 62(2), 245-289.
- Riley, K. & Parker, F. (1986). Anomalous prepositions in relative clauses. *American Speech*, 61(4), 291-306
- Rissanen, M. (1985). Periphrastic do in affirmative statements in Early American English. *Journal of English Linguistics*, 18(2), 163-183.
- Rissanen, M. (1991). Spoken language and the history of do-periphrasis. In D. Kastovsky (ed.), *Historical English syntax*, (321-342). Berlin: De Gruyter Mouton.
- Rissanen, M. (1997). “Candy no witch, Barbados”: Salem witchcraft trials as evidence of early American English. In H. Ramisch, & K. Wynne (eds.), *Language in time and space. Studies in honour of Wolfgang Viereck on the occasion of his 60<sup>th</sup> birthday*, (183-193). Stuttgart: Steiner.
- Rissanen, M. (2003). Salem Witchcraft Papers as evidence of early American English. *English Linguistics*, 20(1), 84-114.
- Rohdenburg, G. (1996). Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics*, 7(2), 149-182.
- Romaine, S. (1980). The relative clause marker in Scots English: Diffusion, complexity, and style as dimensions of syntactic change. *Language in Society*, 9(2), 221-247.
- Romaine, S. (1981). Syntactic complexity, relativization and stylistic levels in Middle Scots. *Folia Linguistica Historica*, 15(21), 71-98.
- Romaine, S. (1982). *Socio-historical linguistics: Its status and methodology*. Cambridge: Cambridge University Press.
- Romaine, S. (2003). Variation in language and gender. In J. Holmes, & M. Meyerhoff (eds.), *The handbook of language and gender*, (98-118). Oxford: Blackwell.

- Romaine, S. (2016). The variationist approach. In M. Kytö, & P. Pahta (eds.) *The Cambridge handbook of English historical linguistics*, (19-36). Cambridge: Cambridge University Press.
- Rosenthal, B., & Adams, G. A. (2009). *Records of the Salem witch-hunt*. Cambridge: Cambridge University Press.
- Royster, J. F. (1915). The “do” auxiliary – 1400 to 1450. *Modern Philology*, 12(7), 449-456.
- Rydén, M. (1966). *Relative constructions in early sixteenth century English, with special reference to Sir Thomas Elyot*. Stockholm: Almqvist & Wiksell.
- Rydén, M. (1983). The emergence of *who* as relativizer. *Studia Linguistica*, 37(2), 126-134.
- Samaha, J. B. (1981). The recognizance in Elizabethan law enforcement. *The American Journal of Legal History*, 25(3), 189-204.
- Samuels, M. L. (1972). *Linguistic evolution: With special reference to English*. Cambridge: Cambridge University Press.
- Sankoff, D. (1988). Sociolinguistics and syntactic variation. In F. J. Newmeyer (ed.), *Language: The socio-cultural context*, (140-161). Cambridge: Cambridge University Press.
- Santelmann, L., Berk, S., Austin, J., Somashekar, S., & Lust, B. (2002). Continuity and development in the acquisition of inversion in yes/no questions: Dissociating movement and inflection. *Journal of Child Language*, 29(4), 813-842.
- Schneider, E. W. (1989). *American earlier Black English*. Alabama: University of Alabama Press.
- Schneider, E. W. (2002). Investigating variation and change in written documents. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (eds.), *The handbook of language variation and change*, (67-96). Oxford: Blackwell.
- Schneider, E. W. (2004). The English dialect heritage of the southern United States. In R. Hickey (ed.), *Legacies of colonial English: Studies in transported dialects*, (262-309). Cambridge: Cambridge University Press.
- Schneider, E. W. (2013). Investigating historical variation and change in written documents. In J.K. Chambers & N. Schilling (eds.) *The handbook of language variation and change*, (57-82). New Jersey: Wiley.
- Schulz, N. (2021). Auxiliary *DO* in Asian Englishes. *World Englishes*, 40(4), 502-518.
- Schütze, C. T. (2004). Synchronic and diachronic microvariation in English *do*. *Lingua*, 114(4), 495-516.
- Smith, C. L. (2018). “Candy no witch in her country”: What one enslaved woman’s testimony during the Salem witch trials can tell us about early American literature. In C. L. Smith, M. P. Grier, & N. R. Jones (eds.), *Early modern black diaspora studies*, (107-134). Cham, Switzerland: Springer International Publishing AG.
- Snyder, W. (2017). On the child’s role in syntactic change. In G. Sengupta, S. Sircar, M. G. Raman, & R. Balusu (eds.), *Perspectives on the architecture and acquisition of syntax*, (235-242). Singapore: Springer Singapore.
- Söderlund, A. (2017). Periphrastic *do* in English witness depositions 1560-1760. *Nordic Journal of English Studies*, 16(1), 244-277.

- Stein, D. (1990). *The semantics of syntactic change: Aspects of the evolution of 'do' in English*. Berlin: De Gruyter Mouton.
- Suárez, C. (2008). Strategies in competition: Pronominal relativizers in the history of English. *English Studies*, 89(3), 339-350.
- Suárez, C. (2012). The consolidation of *þat* as an invariable relativizer in the history of English. *Nordic Journal of English Studies*, 11(1), 79-107.
- Suárez-Gómez, C. (2006). *Relativization in Early English (950–1250): The position of relative clauses*. Bern: Lang.
- Swinburne, N. (2025). The grammaticalization and grammatical meaning of a 'do' support verb in the northern Italian Camuno dialect. *Probus*, 37(1), 1-25.
- Tagliamonte, S. (2002). Variation and change in the British relative marker system. In P. Poussa, (ed.), *Relativisation on the North Sea littoral*, (147–65). Munich: LINCOM Europa.
- Tagliamonte, S. (2011). *Variationist sociolinguistics: Change, observation, interpretation*. New Jersey: Wiley.
- Tagliamonte, S. A. (2013). Comparative sociolinguistics. In J.K. Chambers, P. Trudgill, & N. Schilling Estes, *The handbook of language variation and change*, (128-156). New Jersey: Wiley.
- Tagliamonte, S., Smith, J., and Lawrence, H. (2005). No taming the vernacular! Insights from the relatives in Northern Britain. *Language Variation and Change*, 17(1), 75-112.
- Thompson, R. (1994). *Mobility & migration. East Anglian founders of New England, 1629-1640*. Amherst: University of Massachusetts Press.
- Tieken-Boon van Ostade, I. (1987). *The auxiliary do in eighteenth-century English. A sociohistorical linguistic approach*. Dordrecht: Foris.
- Tiersma, P. M. (2015). Why is legal language so conservative? In L. Solan, J. Ainsworth, R. W. Shuy (eds.), *Speaking of language and law: Conversations on the work of Peter Tiersma*, (11-15). Oxford: Oxford University Press.
- Tottie, G. (1995). The man Ø I love: An analysis of factors favouring zero relatives in written American English. In G. Melchers & B. Warren (eds.) *Studies in Anglistics*, (201–215). Stockholm: Almqvist and Wiksell.
- Tottie, G. (1997). Relatively speaking: Relative marker usage in the British National Corpus. In T. Nevalainen & L. Kahlas-Tarkka (eds) *To explain the present: Studies in the changing English language in honour of Matti Rissanen*, (465–481). Helsinki: Société Néophilologique.
- Tottie, G., & Harvie, D. (2000). It's all relative: Relativization strategies in early African American English. In S. Poplack (ed.) *The English history of African American English*, (198–230). Oxford: Blackwell.
- Tottie, G., & Johansson, C. (2015). Here is an old mastiffe bitch Ø stands barking at mee: Zero subject relativizers in Early Modern English (t)here-constructions. In P. Shaw, B. Erman, G. Melchers, & P. Sundkvist (eds.) *From clerks to corpora: Essays on the English language yesterday and today*, (135-153). Stockholm: Stockholm University Press.

- Tottie, G., and Rey, M. (1997). Relativization strategies in Earlier African American Vernacular English. *Language Variation and Change*, 9(2), 219–247.
- Traugott, E. C. (1992). Syntax. In R. M. Hogg (ed.) *The Cambridge history of the English language*, vol. 1, *The beginnings to 1066*, (168–289). Cambridge: Cambridge University Press.
- Van den Eynden, N. (1996). Aspects of preposition placement in English. In J. Klemola, M. Kytö, & M. Rissanen (eds.), *Speech past and present, studies in English dialectology in memory of Ossi Ihalainen*, (426–446). Frankfurt am Main: Peter Lang.
- Van der Auwera, J., & Genée, I. (2002). English *do*: On the convergence of languages and linguists. *English Language & Linguistics*, 6(2), 283–307.
- Van der Wal, M. J., & Rutten, G. J. (2013). *Touching the past: Studies in the historical sociolinguistics of ego-documents*. Amsterdam: John Benjamins.
- Van Herk, G., & Poplack, S. (2003). Rewriting the past: Bare verbs in the Ottawa Repository of Early African American Correspondence. *Journal of Pidgin & Creole Languages*, 18(2), 231–266.
- Visser, F. T. (1969). *An historical syntax of the English language*. Leiden: E. J. Brill.
- Wagner, S. (2007). Unstressed periphrastic *do* – from Southwest England to Newfoundland? *English World-Wide*, 28(3), 249–278.
- Walker, J. A. (2010). *Variation in linguistic systems*. New York: Routledge.
- Warner, A. (2005). Why *DO* dove: Evidence for register variation in Early Modern English negatives. *Language Variation and Change*, 17(3), 257–280.
- Warner, A. (2017). Periphrastic *DO*. In A. Bergs, & L. J. Brinton (eds.), *The history of English: Early Modern English*, (224–240). Berlin, Germany: De Gruyter Mouton.
- Wasow, T., Jaeger, T. F., and Orr, D. M. (2011). Lexical variation in relativizer frequency. In H. J. Simon & H. Wiese (eds.) *Expecting the unexpected: Exceptions in grammar*, (175–95). Berlin: De Gruyter Mouton.
- Weber, T. (2018). An OT analysis of “*do*”-support across varieties of German. *The Journal of Comparative Germanic Linguistics*, 21(1), 75–129.
- Weinreich, U., Labov, W., & Herzog, M. (1968). *Empirical foundations for a theory of language change*. Texas: University of Texas Press.
- Wischer, I. (2008). What makes a syntactic change stop? On the decline of periphrastic *do* in early modern English affirmative declarative sentences. *Studia Anglica Posnaniensia*, 44, 139–154.
- Wolfram, W. (1991). The linguistic variable: Fact and fantasy. *American Speech*, 66(1), 22–32.
- Woolf, A. (2000). Witchcraft or mycotoxin? The Salem witch trials. *Journal of Toxicology: Clinical Toxicology*, 38(4), 457–460.
- Wright, L. (2017). Language contact. In A. Bergs, & L. J. Brinton (eds.), *The history of English: Early Modern English*, (150–166). Berlin, Germany: De Gruyter Mouton.
- Yamada, A. (2019). Do-support and be-support as defective movements. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, 55(1), 435–449.
- Yáñez-Bouza, N. (2015). *Grammar, rhetoric and usage in English*. Cambridge: Cambridge University



Press.

Zemsky, R. M. (1969). Power, influence, and status: Leadership patterns in the Massachusetts Assembly, 1740-1755. *The William and Mary Quarterly: A Magazine of Early American History*, 26(4), 502-520.

Ziegeler, D. (2004). Reanalysis in the history of *do*: A view from construction grammar. *Cognitive Linguistics*, 15(4), 529-574.