Probabilistic and Multivariate Modelling in Latin Grammar: 
the Participle-Auxiliary Alternation as a Case Study

A thesis submitted to The University of Manchester for the degree of 
Doctor of Philosophy 
in the Faculty of Humanities

2014

James W. R. Brookes

School of Arts, Languages and Cultures
Contents

List of Tables .................................................... 15
List of Figures ................................................... 17
List of Abbreviations ............................................ 18
Abstract .................................................................. 19
Declaration ................................................................ 20
Copyright .................................................................. 21
Acknowledgments ................................................... 22

I Preliminaries ....................................................... 23

1 General Introduction ............................................ 24
   1.1 Background ................................................... 24
   1.2 Research questions and scope ........................... 24
      1.2.1 Construction under consideration as a case study .... 25
      1.2.2 Latin data .............................................. 25
      1.2.3 Information space .................................... 25
   1.3 Brief overview of statistical methods .................... 26
   1.4 Organization of the thesis .................................. 26

2 Outlining the Case Study ......................................... 27
   2.1 Introduction ................................................. 27
   2.2 The morphosyntax of the Latin participle-auxiliary construction ........................................... 27
   2.3 The construction's grammatical variability .......... 28
   2.4 Overview of previous analyses of Latin PAC variability ......................................................... 31
   2.5 Previously claimed language-internal constraints ................................................................. 32
      2.5.1 A free choice? ........................................... 32
      2.5.2 Information structure .................................. 32
      2.5.3 Semantics of the participle ......................... 32
      2.5.4 Participle type .......................................... 33
      2.5.5 Impersonal verbs ...................................... 33
      2.5.6 Number .................................................. 33
      2.5.7 Verbal morphology of the auxiliary .......... 33
      2.5.8 Grammatical properties of the pre-verbal word. ......................................................... 33
   2.6 Problems with previous research ......................... 34
3 The Probabilistic and Multivariate Framework 35
3.1 Introduction ................................................................. 35
3.2 Probability in language .................................................. 35
3.3 Probabilistic and multivariate grammar: what is it? ............... 36
3.4 Corpus-based implementations of probabilistic models .......... 37
  3.4.1 The dative alternation .............................................. 37
  3.4.2 Dutch AVC alternation ............................................. 38
  3.4.3 Interim summary .................................................... 39
3.5 Psycholinguistic evidence .............................................. 39
  3.5.1 Offline evidence .................................................... 39
  3.5.2 Online evidence .................................................... 40
3.6 Conclusions and issues ................................................ 41

II Data and Materials 42

4 Defining and Delimiting the PAC for Operational Purposes 43
4.1 Introduction ................................................................. 43
4.2 Previous definitions and problems .................................... 43
4.3 Definition in the present study ....................................... 45
  4.3.1 Perfect participle + esse ........................................... 45
  4.3.2 Future participle + esse ............................................ 46
  4.3.3 Gerundive + esse ................................................... 46
  4.3.4 Exclusion of other deverbal forms + esse ....................... 47
4.4 Final remarks .............................................................. 48

5 The Choice of Data 49
5.1 Introduction ................................................................. 49
5.2 Excluded datasources .................................................... 49
  5.2.1 Inscriptions ......................................................... 49
  5.2.2 Poetry ................................................................. 50
  5.2.3 Categorical authors ................................................ 50
5.3 Included datasources .................................................... 50
  5.3.1 Cicero ................................................................. 50
  5.3.2 Caesar ............................................................... 51
  5.3.3 Caesar’s continuators .............................................. 51
  5.3.4 Nepos ................................................................. 51
  5.3.5 Varro ................................................................. 51
5.4 Conclusion ................................................................. 52

6 Data Extraction 53
6.1 Electronic text archive ................................................... 53
6.2 Identification of data points ........................................... 53
6.3 Textual issues .............................................................. 54
6.4 Frequency overview ..................................................... 54

III Information Space 58

7 Methodological Preliminaries 59
  7.1 Introduction ............................................................. 59
7.2 The information space ................................................. 59
  7.2.1 Definition .................................................. 59
  7.2.2 Specification of $\Omega$ ......................................... 60
7.3 Operationalizing the predictors ........................................ 60
7.4 Types of variables .................................................. 60
7.5 Statistical modelling ................................................ 61
  7.5.1 Ordinary least squares regression ............................... 62
    7.5.1.1 Least squares ........................................... 62
    7.5.1.2 Statistical inference ..................................... 67
      7.5.1.2.1 Hypothesis testing .................................. 67
      7.5.1.2.2 Error variance ....................................... 67
      7.5.1.2.3 Standard error ....................................... 67
      7.5.1.2.4 Confidence intervals ................................ 67
    7.5.1.3 Regression diagnostics .................................... 68
      7.5.1.3.1 Normality of errors .................................. 68
      7.5.1.3.2 Linearity ............................................. 68
    7.5.1.4 Dummy variable coding for categorical predictors .......... 68
      7.5.1.4.1 Binary variables ..................................... 69
      7.5.1.4.2 Multicategory variables ............................... 70
        7.5.1.4.2.1 Treatment coding ................................ 70
        7.5.1.4.2.2 Sum coding ...................................... 71
  7.5.2 Generalized linear models ....................................... 72
    7.5.2.1 Introduction to binary response models .................... 72
    7.5.2.2 Modelling binary data in linear regression ................ 73
    7.5.2.3 Logit link function ....................................... 75
    7.5.2.4 Maximum Likelihood Estimation ............................ 75
    7.5.2.5 An example .............................................. 76
  7.5.3 Multilevel generalized linear models ............................ 78
    7.5.3.1 Introduction ............................................. 78
    7.5.3.2 Variance components models ................................ 79
    7.5.3.3 Random intercepts and random slope models with a categorical covariate 79
7.6 Explanation of predictor activity ................................... 80
7.7 Summing up .......................................................... 80

8 Multilevel Structure .................................................. 81
  8.1 Introduction ..................................................... 81
  8.2 The GLM null model ............................................... 81
  8.3 Text-based variability ............................................. 82
  8.4 Lemma-based variability .......................................... 84
    8.4.1 Background ................................................ 85
    8.4.2 Reported lexical effects on grammatical variability .......... 85
      8.4.2.1 General literature ..................................... 85
      8.4.2.2 AVC literature ......................................... 85
    8.4.3 Annotation ................................................ 86
    8.4.4 Multilevel modelling ........................................ 86
  8.5 Complex multilevel structure ..................................... 88
  8.6 Summary .......................................................... 89
## 9 Phonology

9.1 Introduction .......................... 91

9.2 Clisis and Prosodic Structure .................. 91

9.2.1 Latin auxiliaries and clisis .................. 92

9.2.2 Cliticood and word order ................. 93

9.2.2.1 General discussion .................. 93

9.2.2.2 Clisis in AVC research ............... 94

9.2.2.3 Interim summary .................. 95

9.2.3 Variable profile and operationalization .......... 96

9.2.3.1 Auxiliary’s cliticood .................. 96

9.2.3.2 Position of the PAC ................. 96

9.2.3.2.1 Motivating a prosodic definition of 2P ........ 96

9.2.3.2.2 Operationalization of prosodic 2P for a dead language .... 100

9.2.4 Statistical modelling .................. 104

9.2.4.1 Auxiliary cliticood .................. 104

9.2.4.2 Position of the PAC .................. 110

9.2.5 Discussion .................. 113

9.3 Rhythmical Optimization .................. 114

9.3.1 Introduction .................. 114

9.3.2 Grammatical variability .................. 114

9.3.3 Rhythmical optimization in AVCs and PACs ........ 114

9.3.4 Operationalization and variable profile .......... 115

9.3.5 Statistical modelling .................. 117

9.3.5.1 Participle/Auxiliary Rhythm ............. 117

9.3.5.2 Auxiliary/Participle Rhythm ............. 120

9.3.6 Discussion .................. 121

9.4 Segmental optimization .................. 123

9.4.1 Introduction .................. 123

9.4.2 Grammatical variability ................. 124

9.4.2.1 General discussion ................. 124

9.4.2.2 Segmental optimization in AVCs ............. 125

9.4.3 Operationalization and variable profile .......... 126

9.4.4 Statistical modelling .................. 127

9.4.4.1 Final segment of the participle and initial segment of the auxiliary .... 127

9.4.4.2 Final segment of the auxiliary and initial segment of the participle .... 129

9.4.4.3 Final segment of the preverbal word and initial segment of the participle .... 131

9.4.4.4 Final segment of the preverbal word and initial segment of the auxiliary .... 134

9.4.5 Discussion .................. 135

9.5 Clausulae .................. 137

9.5.1 Background .................. 137

9.5.2 Previously identified grammatical effects ........ 137

9.5.3 Variable profile and operationalization .......... 138

9.5.4 Statistical modelling .................. 139

9.5.5 Discussion .................. 142

9.6 Final remarks .................. 143
### 10 Grammatical Predictors

10.1 Introduction ................................................. 145

10.2 Verbal morphology of the auxiliary ................................................. 145

10.2.1 Introduction ................................................. 145

10.2.2 Grammatical effects ................................................. 146

10.2.2.1 General effects ................................................. 146

10.2.2.2 AVC/PAC specific effects ................................................. 147

10.2.3 Variable profile and operationalization ................................................. 148

10.2.4 Statistical modelling ................................................. 148

10.2.5 Discussion ................................................. 149

10.3 Morphosyntactic category of the participle ................................................. 151

10.3.1 Introduction ................................................. 151

10.3.2 Grammatical effects in AVCs ................................................. 151

10.3.2.1 Dutch ................................................. 151

10.3.2.2 Latin ................................................. 151

10.3.3 Variable profile and operationalization ................................................. 152

10.3.3.1 Phonomorphology ................................................. 153

10.3.3.2 Morphology ................................................. 154

10.3.3.2.1 Negative prefixation ................................................. 154

10.3.3.2.2 Degree modification ................................................. 154

10.3.3.2.3 Word-formation ................................................. 154

10.3.3.3 Semantics ................................................. 155

10.3.3.3.1 Telicity ................................................. 155

10.3.3.3.2 Specialized meanings ................................................. 155

10.3.3.4 Syntax ................................................. 155

10.3.3.4.1 Complementation ................................................. 155

10.3.3.4.2 Coordination ................................................. 155

10.3.3.4.3 Agentive by-phrase ................................................. 156

10.3.3.5 Summary of diagnostics ................................................. 156

10.3.4 Statistical modelling ................................................. 157

10.4 Prepositional Prefixation of the Participle ................................................. 159

10.4.1 Introduction ................................................. 159

10.4.2 Effect in AVCs ................................................. 160

10.4.3 Operationalization ................................................. 160

10.4.4 Statistical Modelling ................................................. 161

10.4.5 Discussion ................................................. 162

10.5 Impersonality ................................................. 162

10.5.1 Introduction ................................................. 162

10.5.2 Impersonality and PAC serialization ................................................. 163

10.5.3 Operationalization ................................................. 163

10.5.4 Statistical modelling ................................................. 164

10.5.5 Discussion ................................................. 165

10.6 Deponency ................................................. 165

10.6.1 Introduction ................................................. 165

10.6.2 Deponency and PAC serialization ................................................. 166

10.6.3 Operationalization ................................................. 166

10.6.4 Bivariate modelling ................................................. 167

10.6.5 Discussion ................................................. 168

10.7 Conjugational Class ................................................. 169
11.4.2.1 General linguistic importance .............................................. 218
11.4.2.2 Effects in AVCs ................................................................. 219
11.4.3 Variable profile and operationalization ........................................ 220
11.4.3.1 Problems of operationalization ............................................ 220
11.4.3.2 Delimitation of the Discourse Context .................................... 222
11.4.3.3 Definiteness ................................................................. 223
11.4.3.4 Topicality ................................................................. 223
11.4.3.5 Contrastiveness .............................................................. 223
  11.4.3.5.1 Grammatical contexts ................................................... 223
    11.4.3.5.1.1 From-to ...................................................... 224
    11.4.3.5.1.2 Reciprocal .................................................. 224
    11.4.3.5.1.3 Juxtaposition ................................................. 224
    11.4.3.5.1.4 Not-but ..................................................... 224
    11.4.3.5.1.5 Either-or ..................................................... 224
    11.4.3.5.1.6 Not-only-but-also ........................................... 225
    11.4.3.5.1.7 Both-and .................................................... 225
    11.4.3.5.1.8 Clarificatory ................................................ 225
    11.4.3.5.1.9 Comparative ................................................ 225
    11.4.3.5.1.10 Predicative ............................................... 225
  11.4.3.5.2 Lexical contexts ........................................................ 225
  11.4.3.5.3 Pragmatic contexts .................................................. 226
    11.4.3.5.3.1 Polarity ...................................................... 226
    11.4.3.5.3.2 Contrastive ................................................ 226
    11.4.3.5.3.3 Parallel ....................................................... 227
    11.4.3.5.3.4 Superset ...................................................... 227
11.4.4 Statistical Modelling ............................................................ 228
  11.4.4.1 Definiteness ............................................................... 228
  11.4.4.2 Topicality ............................................................... 229
  11.4.4.3 Contrastiveness .......................................................... 230
11.4.5 Discussion ........................................................................ 233
11.5 Summary remarks .................................................................... 235

12 Usage-based Predictors .................................................................. 236
  12.1 Introduction ........................................................................ 236
  12.2 Persistence .......................................................................... 236
    12.2.1 Introduction ................................................................ 236
    12.2.2 Evidence of persistence effects ....................................... 237
      12.2.2.1 General research ................................................... 237
      12.2.2.2 AVC research ....................................................... 238
    12.2.3 Variable profile and operationalization ................................. 239
    12.2.4 Statistical modelling ..................................................... 239
    12.2.5 Discussion .................................................................. 244
  12.3 Frequency ............................................................................ 244
    12.3.1 Introduction ................................................................ 244
    12.3.2 Background .................................................................. 245
      12.3.2.1 General research ................................................... 245
      12.3.2.2 AVC research ....................................................... 245
    12.3.3 Variable profile and operationalization .................................. 246
    12.3.4 Statistical modelling ..................................................... 246
CONTENTS

12.3.4.1 Frequency of the participle ............................................. 246
12.3.4.2 Frequency of the auxiliary ............................................. 249
12.3.5 Discussion ................................................................. 251
12.4 Summary ................................................................. 252

IV Multivariate Modelling .......................................................... 253

13 Methodological Preliminaries ..................................................... 254
  13.1 Introduction ............................................................... 254
  13.2 Motivating random forests ................................................... 254
  13.3 Conditional inference trees ................................................... 255
  13.4 Growing a random forest of trees ......................................... 257
  13.5 Application of random forests in language research ...................... 261

14 Random Forests Model of the PAC Alternation .................................. 262
  14.1 Introduction ................................................................... 262
  14.2 Procedure .................................................................... 262
  14.3 Results and discussion ....................................................... 263
    14.3.1 Prediction accuracy ....................................................... 263
    14.3.2 Variable importance ....................................................... 265
  14.4 Summary .................................................................... 269

V Summing Up ................................................................................. 270

15 Conclusions and Additional Remarks ............................................... 271
  15.1 Summary of findings and implications ...................................... 271
    15.1.1 Predictors of PAC variability .............................................. 271
      15.1.1.1 Multilevel structure ...................................................... 271
      15.1.1.2 Prosodic and phonological predictors .................................... 272
      15.1.1.3 Grammatical predictors .................................................... 272
      15.1.1.4 Semantic and pragmatic predictors ....................................... 272
      15.1.1.5 Usage-based predictors .................................................... 273
    15.1.2 Explanations ............................................................... 273
    15.1.3 Methodological conclusions .............................................. 273
    15.1.4 Operationalization ....................................................... 274
  15.2 Practical applications ......................................................... 274
    15.2.1 Textual criticism ............................................................ 274
    15.2.2 Literary stylistics ........................................................... 275
  15.3 Final remarks .................................................................... 275

A Textual Issues .............................................................................. 295
  Word Count: 94,976
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linguistic abbreviations assumed in the current work</td>
<td>18</td>
</tr>
<tr>
<td>3.1</td>
<td>Probabilistic effect of predictors of the dative alternation (Bresnan et al., 2007)</td>
<td>38</td>
</tr>
<tr>
<td>3.2</td>
<td>Probabilistic effect of predictors on the Dutch participle-auxiliary alternation (De Sutter, 2009)</td>
<td>39</td>
</tr>
<tr>
<td>5.1</td>
<td>Text sources and word counts</td>
<td>52</td>
</tr>
<tr>
<td>6.1</td>
<td>Textual editions used in LIT-A for the texts considered in this study</td>
<td>53</td>
</tr>
<tr>
<td>6.2</td>
<td>Editions used for textual comparison</td>
<td>55</td>
</tr>
<tr>
<td>6.3</td>
<td>The overall distribution of factus est and est factus variants cross-classified by the authors/works under investigation</td>
<td>56</td>
</tr>
<tr>
<td>7.1</td>
<td>Dataset providing an exact linear relationship</td>
<td>63</td>
</tr>
<tr>
<td>7.2</td>
<td>Dataset providing an inexact linear relationship</td>
<td>64</td>
</tr>
<tr>
<td>7.3</td>
<td>Crawley’s (2012) Famous Five for our fictional dataset</td>
<td>66</td>
</tr>
<tr>
<td>7.4</td>
<td>Comparison of a null GLM with various VCMs for the Bresnan et al. (2007) dataset</td>
<td>79</td>
</tr>
<tr>
<td>8.1</td>
<td>Proportions of factus est for various texts</td>
<td>83</td>
</tr>
<tr>
<td>8.2</td>
<td>Intercept adjustments to a null model for text</td>
<td>84</td>
</tr>
<tr>
<td>8.3</td>
<td>Intercept adjustments to a null model for verb lemma</td>
<td>88</td>
</tr>
<tr>
<td>8.4</td>
<td>ANOVA of glmer.vcm.both and glmer.vcm.text</td>
<td>89</td>
</tr>
<tr>
<td>8.5</td>
<td>ANOVA of glmer.vcm.both and glmer.vcm.lemma</td>
<td>90</td>
</tr>
<tr>
<td>9.1</td>
<td>PAC variant cross-classified by the clitichood of the auxiliary</td>
<td>104</td>
</tr>
<tr>
<td>9.2</td>
<td>ANOVA statistics for clitichood models</td>
<td>105</td>
</tr>
<tr>
<td>9.3</td>
<td>ANOVA of baseline VCM and GLMER containing auxiliary clitichood information</td>
<td>106</td>
</tr>
<tr>
<td>9.4</td>
<td>Final GLMER for serialization ~ auxiliary clitichood</td>
<td>106</td>
</tr>
<tr>
<td>9.5</td>
<td>Individual by-text coefficient estimates and by-text clitic information content</td>
<td>108</td>
</tr>
<tr>
<td>9.6</td>
<td>Observations on the verb lemmas animaduerto and conor</td>
<td>110</td>
</tr>
<tr>
<td>9.7</td>
<td>PAC variant cross-classified by the PAC’s position in the prosodic phrase</td>
<td>110</td>
</tr>
<tr>
<td>9.8</td>
<td>ANOVA statistics for position models</td>
<td>111</td>
</tr>
<tr>
<td>9.9</td>
<td>ANOVA of baseline VCM and GLMER containing positional information</td>
<td>111</td>
</tr>
<tr>
<td>9.10</td>
<td>Final GLMER for serialization ~ position</td>
<td>112</td>
</tr>
<tr>
<td>9.11</td>
<td>Confidence intervals for GLMER for serialization ~ position</td>
<td>112</td>
</tr>
<tr>
<td>9.12</td>
<td>GLMER for serialization ~ position</td>
<td>auxiliary = nonclitic</td>
</tr>
<tr>
<td>9.13</td>
<td>GLMER for serialization ~ position</td>
<td>auxiliary = clitic</td>
</tr>
<tr>
<td>9.14</td>
<td>Sapp (2011: Table 4, 23, modified)</td>
<td>115</td>
</tr>
<tr>
<td>9.15</td>
<td>Model AICs of various operationalizations of rhythmical optimization</td>
<td>117</td>
</tr>
<tr>
<td>9.16</td>
<td>Relationship between participle-auxiliary rhythm and serialization</td>
<td>117</td>
</tr>
<tr>
<td>Page</td>
<td>Description</td>
<td>Line</td>
</tr>
<tr>
<td>------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>9.17</td>
<td>ANOVA statistics for various GLMERs containing rhythmic information on the participle/auxiliary boundary</td>
<td>118</td>
</tr>
<tr>
<td>9.18</td>
<td>ANOVA of baseline VCM and selected GLMER containing rhythmic information on the participle/auxiliary boundary</td>
<td>118</td>
</tr>
<tr>
<td>9.19</td>
<td>Final GLMER for serialization ~ containing rhythmic information on the participle/auxiliary boundary</td>
<td>120</td>
</tr>
<tr>
<td>9.20</td>
<td>Relationship between auxiliary-participle rhythm and serialization</td>
<td>120</td>
</tr>
<tr>
<td>9.21</td>
<td>ANOVA statistics for various GLMERs containing rhythmic information on the auxiliary/participle boundary</td>
<td>121</td>
</tr>
<tr>
<td>9.22</td>
<td>ANOVA of baseline VCM and selected GLMER containing rhythmic information on the auxiliary/participle boundary</td>
<td>121</td>
</tr>
<tr>
<td>9.23</td>
<td>Final GLMER for serialization ~ containing rhythmic information on the auxiliary/participle boundary</td>
<td>121</td>
</tr>
<tr>
<td>9.24</td>
<td>Relationship between the preverbal-auxiliary segmental phonology and serialization</td>
<td>127</td>
</tr>
<tr>
<td>9.25</td>
<td>ANOVA statistics for various GLMERs containing final participle/initial auxiliary segmental information</td>
<td>128</td>
</tr>
<tr>
<td>9.26</td>
<td>ANOVA of baseline VCM and selected GLMER containing final participle/initial auxiliary segmental information</td>
<td>129</td>
</tr>
<tr>
<td>9.27</td>
<td>Final GLMER for serialization ~ final participle/initial auxiliary segmental information</td>
<td>129</td>
</tr>
<tr>
<td>9.28</td>
<td>Relationship between the preverbal/participle segmental phonology and serialization</td>
<td>131</td>
</tr>
<tr>
<td>9.29</td>
<td>ANOVA statistics for auxiliary-participle segments</td>
<td>131</td>
</tr>
<tr>
<td>9.30</td>
<td>ANOVA of baseline VCM and selected GLMER containing final auxiliary/initial participle segmental information</td>
<td>131</td>
</tr>
<tr>
<td>9.31</td>
<td>Final GLMER for serialization ~ final auxiliary/initial participle segmental information</td>
<td>132</td>
</tr>
<tr>
<td>9.32</td>
<td>Relationship between the preverbal/participle segmental phonology and serialization</td>
<td>133</td>
</tr>
<tr>
<td>9.33</td>
<td>ANOVA statistics for preverbal-participle segments</td>
<td>133</td>
</tr>
<tr>
<td>9.34</td>
<td>ANOVA of baseline VCM and selected GLMER containing final preverbal/initial participle segmental information</td>
<td>134</td>
</tr>
<tr>
<td>9.35</td>
<td>Final GLMER for serialization ~ final preverbal/initial auxiliary segmental information</td>
<td>134</td>
</tr>
<tr>
<td>9.36</td>
<td>Relationship between the preverbal-auxiliary segmental phonology and serialization</td>
<td>135</td>
</tr>
<tr>
<td>9.37</td>
<td>ANOVA statistics for preverbal-auxiliary segments</td>
<td>135</td>
</tr>
<tr>
<td>9.38</td>
<td>ANOVA of baseline VCM and selected GLMER containing final preverbal/initial auxiliary segmental information</td>
<td>135</td>
</tr>
<tr>
<td>9.39</td>
<td>Final GLMER for serialization ~ final preverbal/initial auxiliary segmental information</td>
<td>135</td>
</tr>
<tr>
<td>9.40</td>
<td>Levels for the clausulae predictor PartAuxMetrics</td>
<td>139</td>
</tr>
<tr>
<td>9.41</td>
<td>Distribution of PAC’s normalized metrical structure and serialization choice</td>
<td>140</td>
</tr>
<tr>
<td>9.42</td>
<td>Constructing a baseline VCM for clausula effects</td>
<td>140</td>
</tr>
<tr>
<td>9.43</td>
<td>ANOVA of baseline VCM and GLMER containing clausula information</td>
<td>141</td>
</tr>
<tr>
<td>9.44</td>
<td>GLMER of serialization ~ clausula in normalized PAC order</td>
<td>141</td>
</tr>
<tr>
<td>9.45</td>
<td>Confidence intervals for the model GLMER of serialization ~ clausula in normalized PAC order</td>
<td>142</td>
</tr>
<tr>
<td>9.46</td>
<td>Summary of phonological and prosodic predictors</td>
<td>144</td>
</tr>
<tr>
<td>10.1</td>
<td>Relationship between the verbal inflection of the auxiliary and PAC serialization</td>
<td>148</td>
</tr>
<tr>
<td>10.2</td>
<td>ANOVA of baseline GLM and selected GLMER containing grammatical number information on the subject NP</td>
<td>148</td>
</tr>
<tr>
<td>10.3</td>
<td>Model parameters of final GLMER for serialization ~ auxiliary’s verbal morphology</td>
<td>149</td>
</tr>
<tr>
<td>10.4</td>
<td>Adjectivehood diagnostics</td>
<td>157</td>
</tr>
<tr>
<td>Table Number</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>10.5</td>
<td>Relationship between participle morphological category and PAC serialization</td>
<td>157</td>
</tr>
<tr>
<td>10.6</td>
<td>ANOVA statistics for various GLMERS containing participle category information</td>
<td>158</td>
</tr>
<tr>
<td>10.7</td>
<td>ANOVA of baseline GLM and selected GLMER containing participle category information</td>
<td>158</td>
</tr>
<tr>
<td>10.8</td>
<td>Model parameters of final GLMER for serialization ~ participle category</td>
<td>158</td>
</tr>
<tr>
<td>10.9</td>
<td>Relationship of prefixation and PAC serialization</td>
<td>161</td>
</tr>
<tr>
<td>10.10</td>
<td>ANOVA statistics for various GLMERS containing prepositional prefix information</td>
<td>161</td>
</tr>
<tr>
<td>10.11</td>
<td>ANOVA of baseline GLM and selected GLMER containing prepositional prefix information</td>
<td>161</td>
</tr>
<tr>
<td>10.12</td>
<td>Final GLMER for serialization ~ prepositional prefix</td>
<td>162</td>
</tr>
<tr>
<td>10.13</td>
<td>Relationship between impersonality and PAC serialization</td>
<td>164</td>
</tr>
<tr>
<td>10.14</td>
<td>ANOVA statistics for various GLMERS containing impersonality information</td>
<td>165</td>
</tr>
<tr>
<td>10.15</td>
<td>ANOVA of baseline GLM and selected GLMER containing impersonality information</td>
<td>165</td>
</tr>
<tr>
<td>10.16</td>
<td>The relationship between deponency and PAC serialization according to Spevak 2006, 2010</td>
<td>166</td>
</tr>
<tr>
<td>10.17</td>
<td>Relationship between deponency and PAC serialization</td>
<td>167</td>
</tr>
<tr>
<td>10.18</td>
<td>ANOVA statistics for various GLMERS containing deponency information</td>
<td>167</td>
</tr>
<tr>
<td>10.19</td>
<td>ANOVA of baseline GLM and selected GLMER containing deponency information</td>
<td>168</td>
</tr>
<tr>
<td>10.20</td>
<td>Final GLMER for serialization ~ deponency</td>
<td>168</td>
</tr>
<tr>
<td>10.21</td>
<td>Relationship between voice and PAC serialization</td>
<td>168</td>
</tr>
<tr>
<td>10.22</td>
<td>Latin Conjunctational Classes</td>
<td>169</td>
</tr>
<tr>
<td>10.23</td>
<td>Relationship of conjunctational class and PAC serialization</td>
<td>170</td>
</tr>
<tr>
<td>10.24</td>
<td>ANOVA statistics for various GLMERS containing conjunctational class information</td>
<td>170</td>
</tr>
<tr>
<td>10.25</td>
<td>ANOVA of baseline GLM and selected GLMER containing conjunctational class information</td>
<td>170</td>
</tr>
<tr>
<td>10.26</td>
<td>Final GLMER for serialization ~ conjunctational class</td>
<td>171</td>
</tr>
<tr>
<td>10.27</td>
<td>Marking of number values (Croft, 2003: 89)</td>
<td>173</td>
</tr>
<tr>
<td>10.28</td>
<td>Frequency of singular nominal forms (Corbett, 2000)</td>
<td>174</td>
</tr>
<tr>
<td>10.29</td>
<td>Relationship between grammatical number and PAC serialization</td>
<td>178</td>
</tr>
<tr>
<td>10.30</td>
<td>ANOVA statistics for various GLMERS containing grammatical number information on the subject NP</td>
<td>178</td>
</tr>
<tr>
<td>10.31</td>
<td>ANOVA of baseline GLM and selected GLMER containing grammatical number information on the subject NP</td>
<td>178</td>
</tr>
<tr>
<td>10.32</td>
<td>Final GLMER for serialization ~ grammatical number</td>
<td>179</td>
</tr>
<tr>
<td>10.33</td>
<td>Gender endings of the participle in the PAC</td>
<td>182</td>
</tr>
<tr>
<td>10.34</td>
<td>Relationship between grammatical gender and PAC serialization</td>
<td>182</td>
</tr>
<tr>
<td>10.35</td>
<td>ANOVA statistics for various GLMERS containing grammatical gender information on the subject NP</td>
<td>183</td>
</tr>
<tr>
<td>10.36</td>
<td>ANOVA of baseline GLM and selected GLMER containing grammatical gender information on the subject NP</td>
<td>183</td>
</tr>
<tr>
<td>10.37</td>
<td>Final GLMER for serialization ~ grammatical gender</td>
<td>184</td>
</tr>
<tr>
<td>10.38</td>
<td>Relationship between grammatical person and PAC serialization</td>
<td>186</td>
</tr>
<tr>
<td>10.39</td>
<td>ANOVA statistics for various GLMERS containing grammatical person information on the subject NP</td>
<td>187</td>
</tr>
<tr>
<td>10.40</td>
<td>ANOVA of baseline GLM and selected GLMER containing grammatical person information on the subject NP</td>
<td>187</td>
</tr>
<tr>
<td>10.41</td>
<td>Relationship between preverbal grammatical category and PAC serialization</td>
<td>189</td>
</tr>
<tr>
<td>10.42</td>
<td>ANOVA of baseline GLM and selected GLMER containing information on the grammatical category of the preverbal expression</td>
<td>190</td>
</tr>
<tr>
<td>10.43</td>
<td>Final GLMER for serialization ~ preverbal grammatical category</td>
<td>191</td>
</tr>
<tr>
<td>10.44</td>
<td>Relationship of clause type and PAC serialization</td>
<td>193</td>
</tr>
<tr>
<td>10.45</td>
<td>ANOVA of baseline GLM and selected GLMER containing clause type information</td>
<td>194</td>
</tr>
<tr>
<td>10.46</td>
<td>Final GLMER for serialization ~ clause type</td>
<td>195</td>
</tr>
</tbody>
</table>
10.47 Summary of effects of grammatical predictors (at p < 0.05), showing alignment which different orders ................................................................. 196

11.1 Eventuality types and features ................................................................. 200
11.2 List of verb classes .............................................................................. 207
11.3 Relationship of eventuality aspect and PAC serialization ......................... 207
11.4 ANOVA statistics for various GLMERs containing eventuality aspect information ......................................................... 208
11.5 ANOVA of baseline VCM and GLMER containing eventuality aspect information ......................................................... 208
11.6 Final GLMER for serialization ∼ eventuality aspect ..................................... 208
11.7 Relationship of verb class and PAC serialization ......................................... 209
11.8 ANOVA of baseline VCM and GLMER containing verb class information .......... 209
11.9 Final GLMER for serialization ∼ verb class ................................................ 210
11.10 Relationship of preverbal animacy and PAC serialization .......................... 215
11.11 Constructing a baseline VCM for animacy effects ......................................... 215
11.12 ANOVA statistics for various GLMERs containing animacy information ........ 215
11.13 ANOVA of baseline VCM and selected GLMER containing animacy information ................................................................. 216
11.14 Relationship of preverbal NP definiteness and PAC serialization ................. 228
11.15 Constructing a baseline VCM for definiteness effects .................................. 228
11.16 ANOVA statistics for various GLMERs containing definiteness information .... 228
11.17 ANOVA of baseline VCM and selected GLMER containing definiteness information ................................................................. 229
11.18 Final GLMER for serialization ∼ preverbal definiteness .................................. 229
11.19 Relationship of preverbal NP topicality and PAC serialization ..................... 230
11.20 ANOVA statistics for various GLMERs containing topicality information ........ 230
11.21 ANOVA of baseline VCM and selected GLMER containing topicality information ................................................................. 230
11.22 Final GLMER for serialization ∼ preverbal topicality .................................. 232
11.23 Relationship of preverbal contrastiveness and PAC serialization ................ 232
11.24 ANOVA of baseline VCM and selected GLMER containing contrastiveness ........ 233
11.25 Final GLMER for serialization ∼ contrastiveness ......................................... 233
11.26 Summary of semantic and pragmatic predictors .......................................... 235

12.1 Example coding of the persistence predictor ............................................. 239
12.2 Distribution of prime and target .............................................................. 239
12.3 Constructing a baseline VCM for persistence effects .................................... 240
12.4 ANOVAs for persistence models ............................................................. 241
12.5 Parameters for persistence model ............................................................ 241
12.6 Model comparisons for persistence and prime lemma models ...................... 242
12.7 Model comparisons for persistence and prime’s auxiliary similarity interaction ................................................................. 243
12.8 ANOVA of baseline VCM and GLMER containing persistence and prime’s auxiliary similarity information ................................................................. 244
12.9 GLMER of persistence and auxiliary similarity ............................................ 244
12.10 Deviance and related statistics ............................................................... 248
12.11 ANOVA of baseline VCM and GLMER containing participle frequency information ................................................................. 248
12.12 GLMER of serialization ∼ participle frequency .......................................... 248
12.13 Deviance statistics for various GLMER models containing auxiliary frequency information ................................................................. 249
12.14 ANOVA of baseline VCM and GLMER containing auxiliary frequency information ................................................................. 249
12.15 GLMER of serialization ∼ auxiliary frequency .......................................... 250
12.16 Cross-tabulation of PAC serialization and combined participle and auxiliary frequency ................................................................. 251

13.1 Observed predictor and permuted predictor of animacy based on Tagliamonte & Baayen (2012) ................................................................. 260
14.1 Random forest confusion matrix for entire sample ................................. 263
14.2 Random forest confusion matrix for out-of-bag observations .................... 263
14.3 Randomly sampled examples of observed values and predicted values ............ 265
List of Figures

6.1 Descending proportions of factus est by authors/works ................................. 56
7.1 Exact linear relationship: \( y = \beta_0 + \beta_1 \) .................................................. 62
7.2 Inexact linear relationship: \( y = \beta_0 + \beta_1 + \epsilon \) ..................................... 65
7.3 Binary variable .................................................................................................. 69
7.4 Plot demonstrating problems with modelling binary in linear regression .......... 73
7.5 Diagnostic plots indicating problems of modelling binary data in OLS regression ... 74
8.1 Multilevel by-text structure characterizing the PAC dataset .............................. 81
8.2 Intercept adjustments to a null model for text ................................................. 84
8.3 Intercept adjustments to a null model for verb lemma ..................................... 89
9.1 Distribution of PAC variants according to the auxiliary’s cliticood .................... 105
9.2 Slope adjustments for auxiliary cliticood by text ............................................. 107
9.3 Slope adjustments for auxiliary cliticood by verb lemma ................................. 107
9.4 Relationship between by-text coefficient estimates (for serialization choice \( \sim \) auxiliary cliticood) and clitic information content ................................................. 109
9.5 Individual by-text coefficients for GLMER rhythmic information on the participle/auxiliary boundary ................................................................. 119
9.6 Individual by-text coefficients for GLMER rhythmic information on the auxiliary/participle boundary ................................................................. 122
9.7 Individual by-text coefficients for GLMER containing final participle/initial auxiliary segmental information ......................................................... 130
9.8 Individual by-lemma coefficients for GLMER containing final auxiliary/initial participle segmental information ......................................................... 132
9.9 Individual by-text coefficients for GLMER containing final preverbal/initial auxiliary segmental information ......................................................... 136
9.10 Barplot of the relationship between normalized metrical structure and serialization choice 140
10.1 Individual by-text coefficients for GLMER containing participle category information 159
10.2 Individual by-text coefficients for GLMER containing conjugation class information 171
10.3 Individual by-text and by-lemma coefficients for GLMER containing grammatical number information ................................................................. 180
10.4 Individual by-lemma coefficients for GLMER containing grammatical gender information 184
11.1 Lambrecht’s (1994: 109) information states ....................................................... 217
11.2 Discourse context for mortuus est (Caes. Gal. 1. 4. 3) .................................... 222
11.3 Contrast matrix ............................................................................................... 227
11.4 Individual by-text coefficients for GLMER containing preverbal topicality information 231
12.1 Kernel density plots for frequency predictors ........................................ 246
12.2 Non-parametric smoothers for frequency of participle ............................. 247
13.1 Illustrative example of the conditional inference tree algorithm .................. 256
13.2 Conditional inference tree for the English dative alternation ..................... 258
14.1 Separation plots for entire sample (top) and out-of-bag observations (bottom) .... 264
14.2 Random forests and GLM variable importance rankings for the predictors considered in this study ................................................................. 266
14.3 Animacy interactions ............................................................................. 267
14.4 Topicality interactions .......................................................................... 268
14.5 Gender interactions ................................................................................ 269
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABL</td>
<td>ablativa</td>
</tr>
<tr>
<td>ACC</td>
<td>accusativa</td>
</tr>
<tr>
<td>ACT</td>
<td>activa</td>
</tr>
<tr>
<td>AUX</td>
<td>auxiliale</td>
</tr>
<tr>
<td>CL</td>
<td>clitic</td>
</tr>
<tr>
<td>COP</td>
<td>copula</td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
</tr>
<tr>
<td>DEM</td>
<td>demonstrativa</td>
</tr>
<tr>
<td>ERG</td>
<td>ergativa</td>
</tr>
<tr>
<td>F</td>
<td>femine</td>
</tr>
<tr>
<td>FUT</td>
<td>future</td>
</tr>
<tr>
<td>GEN</td>
<td>genitiva</td>
</tr>
<tr>
<td>GER</td>
<td>gerundive</td>
</tr>
<tr>
<td>IMPF</td>
<td>imperfective</td>
</tr>
<tr>
<td>IND</td>
<td>indicative</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
</tr>
<tr>
<td>LOC</td>
<td>locative</td>
</tr>
<tr>
<td>M</td>
<td>masculin</td>
</tr>
<tr>
<td>N</td>
<td>neuter</td>
</tr>
<tr>
<td>NEG</td>
<td>negative</td>
</tr>
<tr>
<td>NOM</td>
<td>nominativa</td>
</tr>
<tr>
<td>PAC</td>
<td>participle-auxiliary cluster</td>
</tr>
<tr>
<td>PASS</td>
<td>passive</td>
</tr>
<tr>
<td>PERS</td>
<td>personal</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>PLUPRF</td>
<td>pluperfect</td>
</tr>
<tr>
<td>PRF</td>
<td>perfect</td>
</tr>
<tr>
<td>PRN</td>
<td>pronoun</td>
</tr>
<tr>
<td>PRS</td>
<td>present</td>
</tr>
<tr>
<td>PRT</td>
<td>particle</td>
</tr>
<tr>
<td>PTCPL</td>
<td>participle</td>
</tr>
<tr>
<td>REFL</td>
<td>reflexive</td>
</tr>
<tr>
<td>REL</td>
<td>relativizer</td>
</tr>
<tr>
<td>SBJV</td>
<td>subjunctive</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SPN</td>
<td>supine</td>
</tr>
<tr>
<td>SUPL</td>
<td>superlative</td>
</tr>
<tr>
<td>TAM</td>
<td>tense, aspect, modality</td>
</tr>
<tr>
<td>VOC</td>
<td>vocative</td>
</tr>
</tbody>
</table>

Table 1: Linguistic abbreviations assumed in the current work
Abstract

Recent research has shown that language is sensitive to probabilities and a whole host of multivariate conditioning factors. However, most of the research in this arena centres on the grammar of English, and, as yet, there is no statistical modelling on the grammar of Latin, studies of which have to date been largely philological. The rise in advanced statistical methodologies allows us to capture the underlying structure of the rich datasets which this corpus only language can potentially offer. This thesis intends to remedy this deficit by applying probabilistic and multivariate models to a specific case study, namely the alternation of word order in Latin participle auxiliary clusters (PACS), which alternate between participle-auxiliary order, as in mortuus est ‘dead is’ and est mortuus ‘is dead’

The broad research questions to be explored in this thesis are the following: (i) To what extent are probabilistic models useful and reflective of Latin syntax variation phenomena?, (ii) What are the most useful statistical models to use?, (iii) What types of linguistic variables influence variation, (iv) What theoretical implications and explanations do the statistical models suggest?

Against this backdrop, a dataset of 2409 PAC observations are extracted from Late Republican texts of the first century BC. The dataset is annotated for an “information space” of thirty-three predictor variables from various levels of linguistics: text and lemma-based variability, prosody and phonology, grammar, semantics and pragmatics, and usage-based features such as frequency.

The study exploits such statistical tools as generalized linear models and multilevel generalized linear models for the regression modelling of the binary categorical outcome. However, because of the potential collinearity, and the many predictor terms, amongst other issues, the use of these models to assess the joint effect of all predictors is particularly problematic. As such, the new statistical toolkit of random forests is utilized for evaluating the relative contribution of each predictor.

Overall, it is found that Latin is indeed probabilistic in its grammar, and the conditioning factors that govern it are spread widely throughout the language space. It is also noted that probabilistic models, such as the ones used in this study, have practical applications in traditional areas of philology, including textual criticism and literary stylistics.
Declaration

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright

The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
This thesis would not have been possible without many people. I am very grateful to the supervisory team — David Langslow, John Payne, Andrew Koontz-Garboden, and Benedikt Szmrecsanyi — for stimulating discussion and inspiration. I am especially grateful to David, who has spent what must be untold hours commenting on draft after draft after draft.

I fired off emails, here, there, and everywhere, to loads of people, who were very kind and offered much advice: Gregory Anderson, Brigette Bauer, John Blundell, Miriam Butt, Greville Corbett, Andrew Devine, Matthew Dryer, Nomi Erteschik-Shir, Stefan Th. Gries, Gerd Haverling, Kris Heylen, Graeme Hutcheson, Florian Jaeger, Jeroen Lauwers, Adam Ledgeway, Arne Lohmann, Giuseppe Pezzini, Giampaolo Salvi, Chris Sapp, Eva Schultze-Berndt, Anne Schwarz, Carolin Strobl, Gert de Sutter, Gerhard Tutz, and Malte Zimmerman.

I should very much like to thank Jonathan Bagley, Alexander Donev, and Denis Denisov for permitting me to attend their statistics and probability lectures at the University of Manchester. I have amassed a great deal of statistical knowledge from short courses undertaken at the Cathie Marsh Centre for Census and Social Research (CCSR) at Manchester, particularly the Generalized Linear Modelling courses by Graeme Hutcheson and Nikos Tzavidis, and the multilevel modelling course by Tarani Chandola.

My friends have been a tremendous support: Laura Arman, Mary Begley, Farah Nazir, James Murphy, Danielle Turton, and Fang Yang. I have benefited much from discussions about Latin word order, amongst more mundane things, with Rebekka Ott. I am forever grateful to Laura Arman and James Murphy who read and commented on large parts of the thesis in its final stages, and were swift to say “NO!” when I wanted to put a few more words in. Danielle Turton and James Murphy spent hours correcting horrible blunders in the .bib file.

Douglas Sipple (and his B) deserves an especial mention. He is the best friend ever, and he’s been such a rock over the three-and-a-half years I have known him while being at Manchester. In the final stages of the write-up, he kept my spirits up right up to the end.

Finally, but not least, my parents. They have encouraged me in everything I have done, and it would take another thesis to describe how much they mean to me. It is to them that this thesis is dedicated.

This research was funded by the Mont Follick Scholarship. The thesis itself was typeset in \LaTeX, and the statistical analyses were conducted in R. A debt of gratitude is owed to all the contributors who freely give their time and expertise to support these packages.
Part I

Preliminaries
Chapter 1

General Introduction

1.1 Background

Recent research has shown that language is rich in probabilistic and multivariate phenomena, ranging across its many levels, from phonetics, phonology, morphology and syntax. This is particularly clear when there are competing structural variants.

Dialectal variation between was/were in existential plural constructions in York English is contingent upon language-internal effects such as subject DP constituency, polarity, proximity, alongside demographic effects such as individual speaker effects and age (Tagliamonte & Baayen, 2012). Gries (2001, 2003) demonstrates that the English particle alternation is influenced by a suite of predictors, spanning many levels of linguistic representation. The English dative alternation has been treated by Bresnan, Cueni, Nikitina & Baayen (2007), who find that definiteness, animacy, pronominality, number, person, givenness, persistence, length, and semantic class are effective at discriminating between the NP-dative (Mary gave John the book) and the PP-dative (Mary gave the book to John). De Marneffe, Grimm, Arnon, Kirby & Bresnan (2012) show that these many variables are present in child language too. MacKenzie (2012) shows that auxiliary contraction in English is driven by a plethora of features. Shih, Grafmiller, Futrell & Bresnan (in press) examine the multiple and conflicting factors that bear on genitive choice in English. Finally, Lohmann (2011) shows that the presence of the infinitive marker to after help depends on genre, animacy, constituency realization, inflectional morphology, and length, amongst others.

Advanced statistical techniques lie at the heart of these analyses, as they are able to tease apart the influences of a great deal of predictors.

However, there remains a dearth of analyses approaching the grammar of Latin from such a perspective. Here, studies have for the most part examined single subsets of variables (e.g. discourse pragmatics), and there has been virtually no quantification, in what Gries (2003: 46) describes as “monocausal perspective on an inherently multicausal problem”. This is somewhat surprising given the richness of Latin as a corpus-only language, where large datasets containing many observations and many features are waiting to be constructed. In order to fill this gap in the market, the contributions in the present thesis include applying state-of-the-art statistical techniques in a field that remains largely philological. In addition to applying statistical methodology, it will be shown that these techniques can aid us towards a linguistic interpretation rather than merely being an end in and of themselves.

In sum, in this thesis, I hope to show that such an approach can be profitably applied to the grammar of this language, affording insights on its probabilistic structure and multivariate properties.

1.2 Research questions and scope

Against this backdrop, the following broad research questions will help guide this investigation.
1. To what extent are probabilistic models useful and reflective of Latin syntax variation phenomena?
2. What are the most useful statistical models to use?
3. What types of linguistic variables influence variation?
4. What theoretical implications and explanations do the statistical models suggest?

While it would be nice to study all aspects of grammatical variability in Latin of all periods, to help answer these questions, it is not practically possible to cover everything. I therefore concentrate on specifics I regard as being under the remit of the present investigation, outlined in Section 1.2.1–Section 1.2.3.

### 1.2.1 Construction under consideration as a case study

Like many languages, Latin displays a great amount of grammatical variability at different levels. In terms of morphology, in the third person plural of the perfect tense indicative in the first conjugation, the canonical inflectional morpheme is -auerunt, but it has a rival in -arunt. In syntax, there are realizational and serialization (word order) alternations. For example, the future tense can be realized synthetically (e.g. *portabit* ‘he will carry’) or analytically (e.g. *portaturus est*), and in the NP, the order of the head noun and modifier is not fixed (e.g. *meus amicus* ‘my friend’ vs. *amicus meus*). This thesis cannot possibly do justice to all this constructional variability, and consequently the focus here will be on a relatively neglected type of word order variation — serial variation between the participle and the auxiliary, as demonstrated in (1).

(1) a. mortuus est
die.PTCPL.PRF.M.SG.NOM be.3SG.PRS.IND

b. est mortuus
be.3SG.PRS.IND die.PTCPL.PRF.M.SG.NOM

‘he died’

This construction (I use this term fairly loosely) has received many different labels in the literature. Here, I use the term participle auxiliary cluster/construction (PAC), which — as detailed in Chapter 2 — I envisage as a belonging to auxiliary verb constructions and other types of complex predicates. In sum, this construction will be used to “project out” about Latin grammar, and help answer the questions raised above.

### 1.2.2 Latin data

Latin is quite a vague term, and encompasses all kinds of registers, periods, genres etc. While this additional variation is interesting, it is beyond the scope of the present thesis. The focus will be on Late Republican Latin (90–27 BC) prose, first because, prose is not as susceptible to metrical constraints as verse and second because this period of Latin is associated with the most famous classical writers (e.g. Cicero, Caesar) and ones which readers will be most familiar with. The caveat to the reader is that the statistical analyses presented are applicable only to this time period and modality.

### 1.2.3 Information space

As discussed in various places of this volume, I take grammatical variants to be a function of a set of predictor variables. We term the set of all predictor variables “the information space”, to be denoted \( \Omega \). While this thesis is necessarily limited in terms of its constructional scope and the range of data, a whole range of variables are considered in this study, taken from the literature on various grammatical variation phenomena. They incorporate higher-level predictors, such as verb lemma and text, phonological predictors, grammatical predictors, semantic and pragmatic predictors, and usage-based predictors.
1.3 Brief overview of statistical methods

Statistical analysis is at the core of this thesis, as in the multivariate work mentioned in Section 1.1. A range of statistical techniques will be utilized, including models based on the mathematics of regression — generalized linear models (GLMs) and multilevel generalized linear models (GLMERS) — as well as the more recent tree-based learning technique of random forests, which is more algorithmic in nature.

1.4 Organization of the thesis

The chapters are grouped into five parts. Part I lays the preliminary backdrop for the entire thesis. Chapter 2 spells out in more detail what the PAC case study concerns, and presents an overview of previous scholarship on the subject. Variables previously claimed to bear on the alternation will be discussed. The chapter ends with a review of the problems inherent in the previous scholarship. This serves as a segue into Chapter 3, which discusses more closely aspects of the probabilistic and multivariate enterprise in linguistics, as foil for the present study.

Part II, comprising Chapter 4 through Chapter 6, deals with a presentation of the data on which the case study is based. Chapter 4 presents an operationalization of the response variable, in order to guide data extraction. Chapter 5 motivates the choice of data, arguing why certain choices were made. Chapter 6 reports on how the observations were sourced, and presents some basic frequency information on the overall distribution of the response variable.

Chapter 7 through Chapter 12 constitute the empirical core of the present thesis, Part III, termed here “the information space”. The many predictors that may influence grammatical variability and serialization differences in the Latin PAC are motivated in these chapters, as gleaned from the empirical and theoretical literature in linguistics. Here, too, bivariate statistical models are presented, to give a flavour of how the predictors behave on their own. Chapter 7 sets the methodological ground, presenting the statistical concepts which will feature throughout Part III — the focus being on multilevel generalized linear modelling strategies (“GLMERS”). Chapter 8 discusses how higher-level features in the information space, such as by-text and by-lemma variability, are important to take into account. Chapter 9 examines multiple prosodic/phonological predictors. This chapter also discusses operational issues concerning how to define prosodic phonological phrasing and clisis in Latin. Chapter 10 examines a range of grammatical predictors, where “grammar” is taken to encompass inflectional morphology, morphologically encoded semantic/syntactic concepts, and syntactic features such as clause type and constituency. We then move on to semantic and pragmatic inputs, including animacy, eventuality aspect (“Aktionsart”) and information structure, where we discuss problems of operationalizing these variables. Detailed coding schemes will be presented which attempt to overcome these problems. Chapter 12 explores the effects of the usage-based inputs of frequency and persistence.

Part IV builds on Part III and turns to multivariate modelling. Chapter 13 is the methodological chapter, and discusses problems with extending simple GLMERS to include multiple inputs, particularly considering the vast quantity of predictors in this study. Instead, the main instrument to be exploited is the tree-based classifier random forests, and a detailed overview of how it chooses relevant predictors from amongst many competitors is described. Chapter 14 then applies this method to the information space of the PAC alternation. While identifying the most salient predictors, more particularly it identifies important interaction effects for variables that were deemed otherwise uninfluential in the simple GLMERS. Comparisons of variable importance between predictors deemed influential by the random forests analysis and those predictors deemed (un)influential in standard GLMs analyses will help to facilitate this.

The last unit of the thesis, Part V, comprising Chapter 15, brings the findings of the thesis together, and suggests avenues for further research.
Chapter 2

Outlining the Case Study

2.1 Introduction

This chapter examines aspects of the construction under consideration in more detail than was mentioned in Section 1.2.1. To facilitate an understanding, the following main points will be addressed: an introduction to its morphosyntax/morphosemantics (Section 2.2); its extensive grammatical variability (section Section 2.3); a perspective-by-perspective overview of previous studies of its variability (Section 2.4); a brief overview the language-internal factors claimed by the previous literature to govern it (Section 2.5); and problems with these previous studies (Section 2.6).

2.2 The morphosyntax of the Latin participle-auxiliary construction

This section presents an introduction to the morphosyntax and morphosemantics of the constructional family under review. Our case study concerns bipartite verbal clusters that comprise a deverbal adjective, which I call a “participle”, collocated with a form of the verb *esse* ‘to be’, which I call an “auxiliary”.

With respect to the construction’s morphosyntactic and morphosemantic make-up, the participle contains the lexical information, and is also inflected for grammatical gender (masculine, feminine, neuter), number (singular, plural), and case (nominative or accusative) of the subject; it also houses temporal and diathetic information. Second, the form of the auxiliary verb *esse* ‘to be’ is inflected for person and number of the subject, along with grammatical tense, mood, and finiteness. Some examples of these Latin collocations with various semantic encodings are reported below (not exhaustive).

In (1), there is a perfect passive participle *interfectus* ‘killed’ combined with a present indicative auxiliary *est* ‘is’; the construction roughly translates as a past perfective eventive passive ‘he was killed’.

(1) quo Dion [[interfectus]Ptcpl [est]Aux]PAC
    rel.m.sg.abl Dion.m.sg.nom kill.PTCPL.PRF.M.SG.NOM be.3SG.PRS.IND
    ‘by which Dion was killed’
    (Nep. *Dion* 9. 6)

In (2), there is a future active participle *aperturi* ‘going to reveal’ combined with an imperfect subjunctive form of the verb ‘to be’ *forent*; the construction expresses, in this context, potentiality.

(2) quod inimici eius dissidenti suos
    since enemy.m.pl.nom dem.m.sg.gen quarrel.PTCPL.PRS.M.SG.DAT own.m.pl.acc

*In other examples given in this thesis, the participle and the auxiliary are in bold typeface. In the examples given here, bracketing has also been provided.*
sensus [[aperturi]Ptcpl] [forent]Aux]PAC
feeling.M.PL.ACC reveal.PTCPL.PUT.M.PL.NOM be.3PL.IMPF.SBJV
‘since his enemies would reveal their feelings to one who was at strife with him’
(Nep. Dion 8. 2)

In (3) a gerundive (future passive participle) ponderanda ‘needing to be estimated’ is combined with a present indicative form of the verb ‘to be’ est, and here the construction expresses obligation/necessity.

(3) si per se uirtus sine fortuna [[ponderanda]Ptcpl
if through itself.F.SG.ACC virtue.F.SG.NOM without fortune.F.SG.ABL estimate.GER.F.SG.NOM
[est]Aux]PAC
be.3SG.PRS.IND
‘if virtue is to be estimated without reference to fortune’
(Nep. Thr. 1. 1)

These constructions fit into Anderson’s (2006) broad theoretical definition of “auxiliary-verb constructions” (abbreviated to “avcs”): “a mono-clausal structure minimally consisting of a lexical verb element that contributes lexical content to the construction and an auxiliary verb element that contributes some grammatical or functional content to the construction” (2006: 7), whose “widespread functions . . . cross-linguistically are to encode (or allow for the encoding of) tense, aspect (including inherent/lexical aspect and Aktionsart), and mood categories” (2006: 30).

To distinguish the set of constructions in, e.g., (1)–(3) from other potential types of avcs in Latin, after Anderson’s (2006) definition, I term the former “participle–auxiliary (esse) constructions”, abbreviated henceforth to “PAC”.

2.3 The construction’s grammatical variability

Like most grammatical constructions in Latin, the PAC exhibits striking grammatical variability. When using this construction, a speaker had to make at least three decisions as to its outcome.

First, while it is usual to realize the auxiliary overtly, as in (4a), it has been observed that the auxiliary can be omitted in cases of ellipsis, or “nominal” sentences, such as that in (4b) where the participle profectus takes the full verbal load of the PAC. In other words, a speaker had to choose whether to realize the auxiliary or omit it. I call this strand REALIZATION variation.

Verginius.M.SG.NOM against Aequi.M.PL.ACC depart.PTCPL.PRQ.M.SG.NOM be.3SG.PRS.IND
‘Verginius set out against the Aequi’
(Liv. 2. 63. 4)

If the speaker chose to realize the auxiliary, as in (4a), they then had some further choices to make. They could choose a serialization where the participle profectus precedes the auxiliary est, as in (5a), or one which exhibits the reverse order, with the auxiliary preceding the participle, as in (5b). This second strand I term SERIALIZATION variation.

(5) a. Crassus in fines Vocatium et Tarusatium

2A more operationally explicit definition to drive the derivation of the dataset will be provided in Chapter 4.
3Note that I use “construction” in a fairly informal way. In some approaches to linguistics, e.g. Construction Grammar, “construction” has a more specific meaning that will not concern us in this thesis.
4(4b) is the one actually found in the text.
2.3. THE CONSTRUCTION’S GRAMMATICAL VARIABILITY

[profectus]_{Ptcp} [est]_{Aux} \_PAC
depart.\_Ptcp.\_PRF.\_M.\_SG.\_NOM be.\_3SG.\_PRS.\_IND

‘Crassus set out into the territory of the Vocates and the Tarusates’
(Caes. Gal. 3. 23. 1)

b. Attalus ab Argis Sicyonen
\_Attalus.\_M.\_SG.\_NOM from \_Argos.\_M.\_PL.\_ABL \_Sicyon.\_F.\_SG.\_ACC be.\_3SG.\_PRS.\_IND

[profectus]_{Ptcp} \_PAC
depart.\_Ptcp.\_PRF.\_M.\_SG.\_NOM

Attalus set out from Argos to Sicyon
(Liv. 35. 46. 3)

Corresponding tree diagrams that display this serialization difference structurally are given below in (6) and (7), respectively.\(^5\)

(6)
\[
\begin{array}{c}
\text{Clause} \\
\downarrow \\
\text{NP} \\
\text{Crassus} \\
\text{PP} \\
\text{in fines Vocatium et Tarusatium} \\
\text{PAC} \\
\text{profectus} \quad \text{Aux} \quad \text{est}
\end{array}
\]

(7)
\[
\begin{array}{c}
\text{Clause} \\
\downarrow \\
\text{NP} \\
\text{Attalus} \\
\text{PP} \\
\text{ab Argis} \\
\text{NP} \\
\text{Sicyonen} \\
\text{PAC} \\
\text{est} \quad \text{profectus}
\end{array}
\]

A third decision the speaker had to make was whether to keep the constituents syntactically continuous, as in (5), or to render them discontinuous to whatever degree, as in the examples in (8), where the participle and auxiliary are separated by other words/constituents. I term this final strand of the variation, the PAC’s DEGREE OF DISCONTINUITY.\(^6\) To exemplify, the examples in (5) have a degree of discontinuity of 0, as the participle and auxiliary are contiguous with each other. The examples in (8), by contrast, have a degree of discontinuity \(\geq 1\). More precisely, in (8a), the participle profectus precedes the auxiliary est with discontinuity of another syntactic constituent, the PP per Phocidem, to give a degree of discontinuity of 1; in (8b), the auxiliary erat precedes the participle profectus with discontinuity of an

\(^{5}\)There are different views as to the structure of the Latin clause and its constituents, so I do not want to give the reader the impression that these are accurate representations. Much of the traditional, philological and functional literature assumes a flat or “non-configurational” structure, as in the tree diagrams given here. In the generative-based literature we find at least three views. First is that of Bauer (1995) who assumes that the constituents are “left branching” \([YP] [X_1] [X_2] \ldots [X_n]\) in Latin in contrast to, say, modern day French and English which are “right branching” \([XP] [X_1] [X_2] \ldots [Y_1]\). Second, Devine & Stephens (2006) and Danckaert (2011, 2012) follow Kayne (1994) in assuming a hierarchically configured right branching structure, as in the Universal base, with the surface left branching structure derived by movement operations to satisfy pragmatic constraints and other EPP features; in this respect, the Latin clause is “discourse configurational” in the terms of, e.g., Kiss (1995). Third, Ledgeway (2012) notes that the Latin clause has a fully-fledged complementizer phrase (CP) at the left-periphery, but in all other respects it is flat. I remain agnostic here.

\(^{6}\)This is typically referred to as, variously, hyperbaton, disjunction, Sperrung, Spaltung in the traditional and philological literature on the grammar of Latin. Discussing syntactic discontinuity on the nominal plane, Devine & Stephens (2006: 524) characterize it as “the most distinctively alien feature of Latin word order”.

29
adverb *iam* and PP *ad Caesarem*, with a degree of discontinuity of 2.\(^7\)

(8) a. rex \([\text{profectus}]_{\text{Ptcpl}}\ P_{\text{PAC}} \) \([\text{per} \ Phocidem]_{\text{PP}} \) \([\text{est}]_{\text{Aux}}\ P_{\text{PAC}}\)
   king.M.SG.NOM depart.Ptcpl.PRF.M.SG.NOM through Phocis.F.SG.ACC be.3SG.PRS.IND
   ‘The king set out through Phocis’
   (Liv. 35. 46. 3)

b. Quintus Pilius \([\text{erat}]_{\text{Aux}}\ P_{\text{PAC}}\) \([\text{iam}]_{\text{AdvP}}\) \([\text{ad} \ Caesarem]_{\text{PP}}\)
   Quintus Pilius.M.SG.NOM be.3SG.IMPF.IND already to Caesar.M.SG.ACC
   \([\text{profectus}]_{\text{Ptcpl}}\ P_{\text{PAC}}\)
   depart.Ptcpl.PRF.M.SG.NOM
   ‘Quintus Pilius had already set out to Caesar’
   (Cic. *Att.* 4. 18. 5)

Corresponding flat trees are given in (9) and (10), respectively.

(9)

```
Clause
   NP = rex
   PAC_i
   PP = separator_1 = profectus
   per Phocidem
   Aux = est
   PP = separator_2 = per Phocidem
   PAC_i
```

(10)

```
Clause
   NP = Quintus Pilius
   PAC_i
   AdvP = separator_1 = erat
   iam
   ad Caesarem
   PP = separator_2 = profectus
   PAC_i
```

As to terminology, when discussing the variants in text, mainly for ease of reference, I shall refer to cases of ellipsis as the *factus* variant; to cases where the participle precedes the auxiliary with degree of discontinuity = 0 as the *factus est* variant; to cases where the auxiliary precedes the participle with degree of discontinuity = 0 as the *est factus* variant; to cases where the participle precedes the auxiliary with degree of discontinuity ≥ 1 as the *factus... est* variant; and to cases where the auxiliary precedes the participle with degree of discontinuity ≥ 1 as the *est... factus* variant.\(^8\)

Following the variationist literature, I also assume rough equivalence in meaning between the variants.

---

\(^7\)Operationalized in terms of words, rather than constituents, (8a) exhibits a degree of discontinuity of 2, and (8b) exhibits a degree of discontinuity of 3.

\(^8\)Technically, we might define this as a multivariate outcome vector \(y^T = (y_1, \ldots, y_m)\), as there are three outcome variables of interest (Tutz, 2012: 12, 363–304): (i) realization (a binary variable \(\{0, 1\}\)), (ii) serialization (a binary variable \(\{0, 1\}\)), and (iii) degree of discontinuity (a ratio-scaled variable). Previous research, such as Adams (1994), Spevak (2010), and Brookes (2009), has impressionistically classified the variability in the *pac* as a univariate, multicategory (polytomous) outcome \(k = (k_1, k_2, k_3, k_4, k_5)\) (Tutz, 2012: 207–240), where \(k\) refers to the broad categories (e.g. *factus, factus est, est factus, factus... est, est... factus*), which obviously ignores the alternation’s finer structure, such as degree of discontinuity, thus potentially resulting in information loss. For instance, *factus X est*, with one degree of discontinuity, might not be conditional on the same set of factors as *factus X Y Z est*, with three degrees of discontinuity. In regression terms, the former could be handled by a multivariate response model, the latter by a multinomial logistic response model. Nevertheless, I will use the latter univariate approach to the outcome for expository convenience.
For example, sentences with different PAC serializations such as *rex profectus est* and *rex est profectus* encode the same propositional content ‘the king set out’, thus sharing the same representation, which one might denote *SET-OUT*(king). They are truth-conditionally equivalent: *rex profectus est* is true iff the king set out, and *rex est profectus* is true iff the king set out. Assuming constancy of reference, where (i) *rex* refers to the same king and (ii) *profectus est* and *est profectus* encode the same event, they mutually entail each other. Similar things could be said about the other strands of variation, viz. (i) the choice of realization and (ii) the degree of discontinuity.

As has been said in Section 1.2.1, this investigation concentrates only on the serialization strand of variation, viz. *factus est* vs. *est factus*; however, since the literature typically talks of these featural distinctions together, they will be discussed in the literature reviews that follow.

### 2.4 Overview of previous analyses of Latin PAC variability

The variability identified in the previous section is, according to Spevak (2010: 156), fairly “well known”. It was known to ancient grammarians, e.g. Marius Plotius Sacerdos (late 3rd century AD), as in the following, where he notes the alternation between *licitum est*, a factus est serialization, and *est lictum*, an est factus serialization.

> Quod in primis est uidelicet nostro tempore uitiosum, Tullius non dubitauit uerbo monosyllabo finire structuram, ut ‘ab istius petulantia conservare non lictum est’… [haec compositio demutata faciet] nostri temporis [structuram] sic, …‘ab istius petulantia non est lictum conservare’. (Keil VI. 493.11ff)

‘Cicero did not stop to think about ending a construction with a monosyllabic word, as in *ab istius petulantia conservare non lictum est* ‘[to whom] it was not allowed to keep from that man’s impudence’, a thing which is plainly flawed in our day… . This sequence will undergo change and form the following structure of our age, *ab istius petulantia non est lictum conservare.*’

In more recent times, the variation has been touched on in a broad range of perspectives. It has come under discussion in philological and traditional approaches in monographs and individual studies (e.g. Marouzeau 1908–1909, 1909, 1910, 1938, 1953; Kroll 1921, 1925; Feix 1934; Muldowney 1937; Vogel 1937; Wilkins 1940; Adams 1994; Hoff 1996; Kruschwitz 2004; Brookes 2009); in grammars (e.g. Hofmann et al. 1965; Kühner et al. 1966; Menge et al. 2000); and in commentaries on classical texts (e.g. Pascucci 1965; Gotoff 1993; Dyck 1996; Diouron 1999; Dyck 2003, 2008). It has been studied in a Functional Grammar framework derived from Dik (1997), e.g. Spevak (2006, 2010), where it is used as part of an argument that seeks to stress the importance of pragmatics on Latin grammatical variation. The construction has been of some interest in structural approaches. In the spirit of Chomsky’s (1957) early work, it has been a topic for discussion in Ostaﬁn (1986) and in Salvi (2004). It has been mentioned in discourse- configurational generative approaches, following e.g. Kiss (1995) (Devine & Stephens, 2006; Danckaert, 2011, 2012). Generative inspired typological work has mentioned it in passing too (e.g. Warner 1980, Bauer 1995).

Despite these analyses, what drives the variation has nevertheless been characterized in recent work as being perplexing and hard to pin down. Devine & Stephens (2006: 179) observe on the one hand that “[t]he location of the auxiliary is a complex issue… . It is subject to a number of quite subtle semantic and syntactic conditions, which can sometimes be conflicting and pitted against each other in an optimality calculus according to the number and importance of those favouring postverbal position versus those favouring some preverbal position”. Spevak (2010: 154), on the other, puts it more succinctly: “[the variability] is difficult to explain”. I will argue later that this difficulty is because of how it has been approached in the past. Before I do so, I examine the factors claimed to govern it (Section 2.5).
2.5 Previously claimed language-internal constraints

We shall now briefly zoom in on what language-internal factors the previous literature mentioned in Section 2.4 has claimed impinge on the variability. To simplify things somewhat, we shall concentrate mainly on the serialization and discontinuity strands of the variation, and pass over ellipsis. Also, since many of these constraints are discussed in more depth later, because they feature as predictors in the present study, only brief remarks on a non-exhaustive list of constraints will be offered here.

2.5.1 A free choice?

As Marouzeau (1910: 39) notes, early studies of this alternation were not really interested in defining the variables that it is conditional on: “La valeur de cette inversion...a pu paraître négligeable”. Ritschl (1879), for instance, is one of those to be singled for criticism here, for judging, like many others of his age, “l’ordre des mots selon des préférences personnelles fondées sur un vague sentiment d’élégance ou d’harmonie”. Thus, the serialization choice was largely regarded as a free one without any clear determinants.

2.5.2 Information structure

Like many alternations in Latin grammar, one factor that has been widely claimed to exert an effect on the construction’s variability is the information structural notion of “focus” — i.e. novelty/contrastiveness in the discourse context.9 In this regard, it has been argued that the auxiliary is sensitive to the locus of the focus in the clause.

If the event encoded by the participle contributes new information and/or is contextually contrastive, the focus is on the participle, and, it has been claimed, the auxiliary follows it either directly (factus est) or with discontinuity (factus...est) (e.g. Marouzeau, 1909, 1910, 1938; Muldowney, 1937; Wilkins, 1940; Devine & Stephens, 2006; Spevak, 2010). Contrastiveness on the participle has been argued to be the typical cause of discontinuities of the type factus...est (cf. i.a. Adams 1994; Brookes 2009).

The est factus and est...factus variants have been argued to be used where the eventuality encoded by the verb is presupposed or information easily accommodated into the presupposition, and the preverbal constituent is under narrow focus (cf. e.g. Walden 1896; Marouzeau 1909, 1910, 1938; Devine & Stephens 2006; Spevak 2010).

While informational determinants have been argued to be influential in much of the literature on this alternation, others disagree. For instance, Skutsch (1912) and Kroll (1921) argue that metrical constraints may take priority, and Hoff (1996) observes “Cependant, notre examen du corpus césarien nous amène à constater que ces cas, indiscutables, ne représenter en fait qu’une proportion réduite des occurrences de [est factus]” (1996: 375) and “Ces faits ne s’expliquent pas par un besoin de “souligner” tel ou tel constituant” (1996: 376).

2.5.3 Semantics of the participle

The decompositional semantics of the (perfect) participle have attracted attention. It has been asserted by Devine & Stephens (2006) that the factus est serialization is positively associated with eventualities with an event component, and dispreferences associating with eventualities with only a state component (2006: 180): “for the auxiliary to be postverbal [i.e. factus est] the verb...is almost always eventive...rather than a stative property predicated of the subject”. One example they cite later which may be seen as support for this assertion is the case of profectus est ‘he departed, set out’ (from proficiscor ‘depart,
set out’), an eventive verb which “almost always has the auxiliary after the participle in main clauses” (2006: 188).

2.5.4 Participle type

It has been suggested by Wilkins (1940) and Brookes (2009) that gerundives, that is participles with a morphologically encoded obligative/necessitive component prefer, est (…) factus serializations.

2.5.5 Impersonal verbs

In Latin, verbs can either be personal in the sense that they can take a subject, e.g. bellum gestum est ‘war was waged’, or impersonal in that they take no subject, e.g. pugnatum est ‘lit. it was fought = there was fighting/fighting took place’. Based on a selection of data from Caesar, Devine & Stephens (2006: 181) assert that impersonal verbs “normally” occur with a factus est serialization. The verb pugno ‘to fight’, for example, “occurs predominantly” with the factus est variant in the impersonal passive (2006: 188).

2.5.6 Number

Steele (1913) observes that with future participle + infinitival esse, there is a correlation between number and choice of serialization. Specifically, it is claimed that plural forms are associated the est factus variant and singular forms are associated with the factus est variant.

2.5.7 Verbal morphology of the auxiliary

Möbitz (1923: 24) identifies an effect of the auxiliary’s grammatical tense form on the choice of variant: “Interessant erscheint aber vor allem die Beobachtung, daß in den Metamorphosen beim Stellungstypus [est factus], die Formen von esse fast überall vom Perfektstamm, meistens Plusquamperfekt-Formen, abgeleitet sind; ich habe nur 10 Belege für [present stem esse + participle] feststellen können gegen 28 Beispiele für [perfect stem fuisse + participle].” On the next page this is followed by: “Die reguläre Verbindung [participle + present stem esse] überwiegt bei Apuleius”. The same effect is also claimed to be in operation in the works of St. Augustine of Hippo/Augustine. Muldowney (1937: 133f.) states: “an examination of the instances of [est factus] shows several instances with the perfect tense of the copula”. Similar claims are to be found in Wilkins (1940) and in Marouzeau (1910, 1938).

2.5.8 Grammatical properties of the pre-verbal word.

It has been argued that certain grammatical categories “attract” the auxiliary. One prominently mentioned category is the effect of the negative adverb (cf. e.g. Marouzeau 1910; Lease 1919; Kroll 1921; Müller 1924; Hofmann et al. 1965; Adams 1994; Devine & Stephens 2006). Aside from the negative, the auxiliary has also been argued to be attracted to other types of constituents in a similar way. These are temporal adverbs (Adams, 1994: 38f.), demonstratives (Adams, 1994: 24–28, 37), modifiers involved in comparison (Adams, 1994: 24, 36), quantifiers, numerals, and measuring expressions (Adams 1994: 19f., 35–36; Devine & Stephens 2006: 182, 191), relative pronouns (Adams 1994: 44–53; Devine & Stephens 2006: 191), degree expressions (Adams 1994: 24, 38; Devine & Stephens 2006: 182, 191), subordinators/complementizers (Vogel, 1937; Devine & Stephens, 2006), and interrogatives (Devine & Stephens, 2006: 191).
2.6 Problems with previous research

There are a number of problems with previous studies on the PAC alternation, which to some extent are unfortunately characteristic of studies of Latin grammatical variability more generally.

First, there has been some attempt to rank constraints according to a hierarchy (cf. the quotation from Devine & Stephens (2006) in Section 2.4 above), but unfortunately this is done largely impressionistically, and there has been no quantification (not even straightforward $\chi^2$-tests of independence, for instance).

A second problem, which will be discussed in Part II and Part III in more depth, concerns the lack of clarity with regard to the choice of data and the operationalization of variables. For instance, the choice of data is unmotivated (why was a specific data source chosen, why not another?), and criteria for identifying values/levels of the variables up for consideration, particularly semantic and information structural ones, are not given, and when they are, it is difficult to actually go through a text and apply them in practice.

Third, some potentially influential factors, which have been mentioned in psycholinguistics literature (e.g., effects of complexity, frequency, and persistence), have all but been ignored. Might it not be important to take these into account?

Given the introspectiveness and lack of quantitative rigour in previous approaches, it is of no surprise to find statements such as those in Devine & Stephens (2006) and in Spevak (2010) with regard to the difficulty in pinning down what is important here. There is hence a clear need to integrate the study of this alternation — and, I would argue, Latin grammatical variability more generally — into a probabilistic grammar framework that is informed by a multivariate set of predictive features. In the next chapter, we shall examine the advantages of such an approach as applied in previous probabilistic studies of language.
Chapter 3

The Probabilistic and Multivariate Framework

3.1 Introduction

This chapter discusses the broad conceptual framework of the present study as backdrop for what follows — probabilistic and multivariate grammar, which has its roots in probabilistic approaches to formal grammar (e.g. Booth 1969) and in early sociovariationist theory (e.g. Labov 1969). Before we begin exploiting such models, it is necessary to explain why they are a plausible representation of the linguistic system, since, for instance, there is little point in adopting them if they do not have any grounding in any psychological reality. In this spirit, the present chapter aims to introduce the notion of how probabilistic models can enhance our understanding of grammar and how a language user’s knowledge of grammar is implicitly probabilistic and sensitive to a suite of context-dependent and context-independent cues.

The structure of this chapter is as follows. First, in section 3.2 I review the underlying motivations for assuming a probabilistic language faculty (following heavily Bod et al. 2003). Then in section 3.3 I define how I conceive of a probabilistic and multivariate grammar for the purpose of this thesis. Section 3.4 reports several corpus studies which have implemented probabilistic approaches. In section 3.5 I will examine some experimental evidence — from comprehension and production — that language users really are aware of the probabilistic weights of multiple factors operating in syntax. Finally, section 3.6 will establish what we have learned so far about probabilistic and multivariate grammars.

3.2 Probability in language

The notion prevalent in many mainstream theories of grammar, such as Chomskyan generative linguistics, is that implicit knowledge of language is categorical. There is no room for gradience or for numbers, and if there is variation, it is attributed to performance (extra-linguistic factors). Thus, in the words of Bod et al. (2003: 1): others consider that “[p]erformance may be full of fuzziness, gradience, and continua, but linguistic competence is not”. Indeed, Chomsky (1969: 57) once asserted that “that the notion of ‘probability of a sentence’ is an entirely useless one, under any known interpretation of this term”. However, there is now a substantial amount of evidence that competence — that is, implicit knowledge — of language is probabilistic. Bod et al. (2003: 2–7) point to the following reasons for assuming an inherently probabilistic language faculty, citing evidence from numerous linguistic and psycholinguistic studies.

Cognition It has been shown that from vision to information processing the human mind is a strongly probabilistic learner (see the contributions in e.g. Chater & Manning 2006; Chater & Oaksford 2008).

Inherent variability Language generates lots of variability, and language users seem to know how
to exploit this variation in different communicative settings. If they did not know how to exploit it, then it surely cannot just be down to performance and errors of speech.

**Gradience in well-formedness** It is often assumed in categorical linguistics that grammatical judgements are (usually) binary: either something is well-formed or it is ill-formed, and the sentence being judged gets a ‘*’ put before it or does not. However, there is robust evidence that there is no clear distinction between a sentence that is grammatical and one that is ungrammatical: accordingly, Bod et al. (2003: 4) submit that there “is a cline of well-formedness, wherein some constructions are highly preferred, others are used less frequently, and some are used not at all”. With respect to the well-formedness judgements of new words, studies have shown that these can be predicted as a function of their subparts, such that phonotactic knowledge is probabilistic.

**Frequency** Frequency has a striking impact on language processes: frequent words and frequent structures are accessed faster, are shorter, and undergo phonetic reduction. Thus, language users are clearly storing information about the distribution of linguistic units, based on their prior experience of those units.

**Learning** According to the “poverty of the stimulus” argument, negative evidence is not available to children, and thus grammars must be innate. However, it’s been shown that probabilistic grammars are learnable from positive evidence alone. In other words, children track probabilities of the distribution of particular linguistic units.

**Gradience in functional and category membership** Some adjuncts are clearly adjuncts and some arguments are clearly arguments; but often it is unclear, such that the category “argument-adjunct” has been proposed. Similarly, some participles are clearly adjectival and others clearly verbal; however, diagnostic tests that are used to distinguish them can be problematic. Further, according to Sleeman (2011), there is a cline between adjectival participles and verbal participles.

Thus there is a substantial amount of evidence that the language faculty is probabilistic and competence and performance inform each other. Speakers, in both comprehension and production, use probabilistic inference in language. It is therefore highly likely that the language faculty is inherently probabilistic like other cognitive systems. In sum, it seems plausible to assume that mathematically modelling language probabilistically has clear support at all levels of language representation, following Bod et al.’s (2003: 10) assumption.

### 3.3 Probabilistic and multivariate grammar: what is it?

In this thesis, we focus on one approach to probabilistic linguistics: probabilistic and multivariate grammars (i.e. morphosyntax). I define them in the following way:

- There exists a set of symbolic, structural and/or functional units (e.g. N, V, P, subject, object, focus, topic, etc.);
- The grammar has combinational rules, such as $S \rightarrow NP \ VP$, where the language is configurational in the sense of having hierarchically structured projections or where the language is non-configurational or “flat”;
- Language users have an idea of the probabilities with which these structural units combine. For instance, a VP that contains a V and an NP is quite probable ($p \approx 1$), whereas a VP that contains two NPs and no V is quite impossible ($p \approx 0$).
• Also, language users have some idea of ‘choice’ in language — they are aware of what structurally alternates with what. So in English, language users have the knowledge that, for instance, in particle verb constructions, they can put the particle immediate after the verb (e.g. *John turned on the light*) or after the direct object (e.g. *John turned the light on*).

• Language users have knowledge of an “information space”, denoted $\Omega$, that is the set of all variables that bear on syntax choice.

• Language users have a implicit knowledge of the probabilities with which variables in the information space exert influence on different syntactic choices. We focus on this latter aspect in the sections to follow.

3.4 Corpus-based implementations of probabilistic models

We will now review several examples of how probabilistic models have been used to capture the complex structure of morphosyntactic variability. As case studies, we draw on the probabilistic literature of (i) the dative alternation in English and (iii) participle-auxiliary order in Dutch, the former being likely to be familiar to the reader, the latter being of a similar constructional type to that of the present study’s.

3.4.1 The dative alternation

The dative alternation can be characterized as the choice between two syntactic structures, exemplified in (1).

\begin{itemize}
  \item a. Jack sold [the car] [to Jill]
  \item b. Jack sold [Jill] [the car]
\end{itemize}

When constructing such a sentence with a dative component, the speaker starts with *Jack sold* and is then left with an option of how to carry on the sentence. In (1a), we have a VP which displays the internal structure of (i) an NP that functions as the direct object theme and (ii) a PP that functions as the dative recipient — this is the prepositional dative construction. In (1b), in contrast, we have a VP which displays the internal structure of two NP objects: the first NP functions as dative recipient (the indirect object) and the second NP functions as the direct object — this is the double object construction. For the probabilistic modeller, the question of interest is what conditions the choice between the two realizations? In other words, how does the speaker gauge whether to go for one or the other structure? As Bresnan et al. (2007) have demonstrated, the choice is subject to a multivariate information space and each predictor exerts an influence on the alternation in a gradient and probabilistic way.

Bresnan et al. (2007)’s study utilizes a dataset comprising 2360 datapoints extracted from the Switchboard corpus of spoken American English (recorded telephone conversations).

The observations are coded for a predictor space of 15 input variables, which are reported as being of influence to syntax in the literature: the fixed effects include (1) semantic class of the verb, (2) accessibility of recipient, (3) accessibility of theme, (4) pronominality of recipient, (5) pronominality of theme, (6) definiteness of recipient, (7) definiteness of theme, (8) animacy of recipient, (9) person of recipient, (10) number of recipient, (11) number of theme, (12) concreteness of theme, (13) structural parallelism in dialogue, (14) length difference (log scale); the random effect (15) is verb sense.

A mixed-effects logistic regression model is constructed with respect to the above-mentioned dataset. Thus fitted, all parameters are significant, except the number of recipient. On unseen data, this model accurately predicts 94% of cases. Simplifying somewhat (i.e. ignoring the fine-detail of effect sizes), they determine a probabilistic structure of the alternation along the lines of table 3.1. When controlling for other variables, each variable is shown to be an influential impingement on the constructional choice.

\footnote{Note that $\Omega$ also refers to the sample or outcome space.}
CHAPTER 3. THE PROBABILISTIC AND MULTIVARIATE FRAMEWORK

Table 3.1: Probabilistic effect of predictors of the dative alternation (Bresnan et al., 2007)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>VP[NP NP] more probable → VP[NP PP]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Givenness</td>
<td>given recipient &gt; non-given recipient</td>
</tr>
<tr>
<td></td>
<td>non-given theme &gt; given theme</td>
</tr>
<tr>
<td>Pronominality</td>
<td>pronominal recipient &gt; non-pronominal recipient</td>
</tr>
<tr>
<td></td>
<td>non-pronominal theme &gt; pronominal theme</td>
</tr>
<tr>
<td>Definiteness</td>
<td>definite recipient &gt; indefinite recipient</td>
</tr>
<tr>
<td></td>
<td>indefinite theme &gt; definite theme</td>
</tr>
<tr>
<td>Animacy of recipient</td>
<td>animate &gt; inanimate</td>
</tr>
<tr>
<td>Person of recipient</td>
<td>local (1, 2) &gt; non-local (3)</td>
</tr>
<tr>
<td>Number</td>
<td>plural recipient &gt; singular recipient [NS]</td>
</tr>
<tr>
<td></td>
<td>singular theme &gt; plural theme</td>
</tr>
<tr>
<td>Concreteness of theme</td>
<td>non-concrete &gt; concrete</td>
</tr>
</tbody>
</table>

What this means is that if the recipient is given, pronominal, definite, animate, local, and the theme is non-given, non-pronominal, indefinite, singular, and non-concrete, then the likelihood that the VP[NP NP] structure will be chosen in very high indeed. By contrast, if the recipient is non-given, non-pronominal, indefinite, non-local, and the theme is given, pronominal, definite, plural and concrete, then the likelihood that the VP[NP NP] structure will be chosen by the speaker is very low indeed, and accordingly they will opt for the VP[NP PP] structure. However, there is a lot of gradient ground, and a different mix of levels of the predictors will yield different probabilities for one or the other variant.

3.4.2 Dutch AVC alternation

Much like the Latin PAC alternation, Dutch AVCSs can vary in their internal serialization. In subordinate clauses, the auxiliary can precede the participle, as in (2a), or alternatively the participle can precede the auxiliary, as in (2b).

(2) a. dat ze een goed werkstuk heeft afgeleverd
    that she a good paper has submitted

   b. dat ze een goed werkstuk afgeleverd heeft
    that she a good paper submitted has
    ‘that she has submitted a good paper’

(Examples from De Sutter 2009: 225–226)

This alternation has been examined by De Sutter (2005, 2009). For his data, De Sutter makes use of the CONDIV corpus, focusing on the newspaper genre of Belgian Dutch written between August and December 1996 (to control for possible sociolinguistic effects). 2390 observations are extracted, of which 66.99% are in the auxiliary-participle order. The dataset is then hand-coded for 10 input variables in the predictor space: (1) distance between previous word accent and participle accent (no. of unaccented syllables), (2) distance between following word accent and participle accent, (3) morphological structure of the participle, (4) grammatical relationship between the extrapolated constituent and its head, (5) length of the middlefield, (6) information value of the last preverbal word, (7) inheritance of the last preverbal content word, (8) type of finite verb/categorial status of the participle, (9) syntactic persistence, (10) frequency of the participle. A multiple logistic regression model is then estimated with respect to the dataset, which significantly deviates from the null model and has a high index of concordance $C = 0.8$ thus indicating a good model fit. I simplify De Sutter’s quantitative results in table 3.2.

What this means is that: if the auxiliary is an auxiliary of time, the participle has a separable prefix, the middlefield is long, the information value of the preverbal word is high, there is a previous AUX-PTCP structure in the preceding discourse, the postverbal field is empty, and the frequency of the participle is high, then the probability of a language user opting for the AUX-PTCP variant is very high.

38
3.5. PSYCHOLINGUISTIC EVIDENCE

Indeed; conversely, if the auxiliary is coupled with an adjectival participle, the prefix on the participle is inseparable, the middlefield is shorter, the previous verb cluster is in the ptcpl-aux order, a complement of the main verb follows the cluster, and the frequency of the participle is low, then the probability of getting an aux-ptcpl order is thus very low, and thus the speaker would opt for the ptcpl-aux variant.

Because it is structurally similar to the PAC alternation, we will refer to it at various points in the thesis.

3.4.3 Interim summary

The above two studies demonstrate (i) the multivariate space of conditioning factors and (ii) probabilistic statements over those factors to which alternations are sensitive. Many other recent synchronic studies of syntactic alternations could be reported which show similar stochastic characteristics (Gries, 2001; Szmrecsanyi, 2006; Wulff, 2003; Hilpert, 2008; Tagliamonte & Baayen, 2012; Hoffmann, 2011; Wiechmann & Kerz, 2013).

3.5 Psycholinguistic evidence

In the previous section we have seen evidence from corpus-based studies (in terms of language production). However, one question that might be asked is this: are probabilistic grammars grounded in psychological reality? To answer this question, I now report offline and online psycholinguistic evidence/cognitive validation for such grammatical models as derived from corpora.

3.5.1 Offline evidence

Bresnan (2007) was interested in shedding light on this issue from an offline perspective, utilizing information from the corpus model of Bresnan et al. (2007) on the dative alternation. She hypothesizes, pre-empirically, that, if the probabilistic model estimated from the corpus is an accurate reflection of the syntactic knowledge of English language users, then those same language users should be able to predict with a high degree of accuracy the actual choices made by the original speakers of the corpus dataset. Thus, if the corpus model assigns a high probability to certain syntactic structure, given a multivariate information space of predictor variables, then English language users should assign a high probability of naturalness to that certain syntactic structure, given the same context and the multivariate space of predictor variables. If the corpus model assigns a low probability to a certain syntactic structure, given a multivariate information space of predictor variables, then English language users should assign a low probability of naturalness to that certain syntactic structure, given the same context and the multivariate space of predictor variables. Lastly, if the corpus model assigns an equal probability to both syntactic structures (i.e. it fails to discriminate what distinguishes them), given a multivariate information space.

Table 3.2: Probabilistic effect of predictors on the Dutch participle-auxiliary alternation (De Sutter, 2009)
of predictor variables, then English language users should also assign them more or less equal naturalness ratings.

With respect to the materials, six dative sentences are randomly sampled from five probability bins: those with a very low probability of the VP[NP PP] outcome; those with a low probability of the VP[NP PP] outcome; those with a medium probability of the VP[NP PP] outcome (i.e. where the choice is all but indeterminate, given the model predictors); those with a high probability of the VP[NP PP] outcome; and finally those with a very high probability of the VP[NP PP] outcome. As subjects, Bresnan takes 19 Stanford undergraduates, both male and female, who reported that they were monolingual and had not taken syntax courses. The subjects are then presented with the context of a dative sentence, followed by both the original variant (whichever it was) and the alternative variant. These subjects are asked to rate both of the variants in terms of their “naturalness”: a score of 100 is to be distributed over both sentences. Thus, if the subject assigns 60 to one variant, they should then assign 40 to the other; if the subject assigns a rating of 98 to one of the variants, then they should assign a rating of 2 to the other; and so on.

There are two important findings that emerge from Bresnan’s experiment. First, the ratings are roughly linearly associated with the corpus probabilities; for each subject, as the corpus probability tends towards 1, then the mean score tends towards 100. In other words, there is a positive correlation between the corpus probability and the mean score. Second, a linear mixed effects regression is constructed to model the ratings of the subjects as a function of the same set of predictor variables that went into the corpus model. What she finds is that the regression coefficients, in terms of their magnitudes and their effect sizes, are in the same direction (positive/negative) as the corpus logistic regression model, apart from one of the predictors — the animacy of the recipient. In the corpus model, inanimate recipients prefer to be realized in the VP[NP PP] structure; in the ratings model, subjects rated inanimate recipients as better in the VP[NP NP] structure. However, as Bresnan notes, this may be due to data sparsity (there are only two inanimate recipients in the dataset, and these occur in sentences that are heavily polarized to the VP[NP NP] structure, anyway).

3.5.2 Online evidence

Tily, Gahl, Arnon, Snider, Kothari & Bresnan (2009) offer online evidence from language production, also utilizing information from the corpus model of Bresnan et al. (2007) on the dative alternation. Tily et al. (2009) hypothesize that syntactic probabilities (i.e. the probabilities with which speakers expect a variant to occur, given a suite of variables) bear on fluency in pronunciation. Tily et al. (2009: 151) hypothesize that “If pronunciation variation is a sufficiently sensitive reflection of the multiple probabilistic cues predicting the choice between syntactic structures, then it can help show whether the human language production system does indeed rely on the full range of available cues”. To evaluate this hypothesis, they examine whether the syntactic probabilities as derived from the Bresnan et al. 2007 corpus model can predict (i) the duration of the word to in the VP[NP PP] structure (the idea being that durations are shortened in easily accessible speech), and (ii) the number of disfluencies (the idea being that the number of disfluencies correlates with problems in planning) in the original Switchboard corpus. Controlling for other potentially confounding variables, they first examine, with a linear mixed effects model, the duration of the word to modelled as a function of syntactic probability. They find that higher probability outcomes of the VP[NP PP] structure do significantly predict that the duration of to will be shorter. They then construct a mixed effects logistic regression model, the outcome of interest being presence vs. absence of disfluencies, again modelled as a function of syntactic probability. Again, their results confirm that, if the dative structure is more probable, then there tend to be fewer disfluencies; conversely, if the dative structure is less probable, then disfluencies tend to be present. They conclude firmly that “[t]he present work supports the view that many factors jointly shape speakers’ probabilistic knowledge of language” (Tily et al., 2009: 162).
3.6 CONCLUSIONS AND ISSUES

In summary, there is strong evidence from these online and offline experiments that language users have somehow internalized the probability of one structure occurring over others, given a richly specified information space.

3.6 Conclusions and issues

The purpose of this chapter was to present evidence from case studies in the previous literature that language is strongly multivariate, in being sensitive to a host of variables, and probabilistic, in that the effect these variables have ranges over \([0, 1]\). However, there are several points worth noting. First, much of the discussion is centred around modern Germanic, mainly English, morphosyntactic alternations. It is therefore unclear to what extent such probabilistic models are appropriate for (i) corpus-only languages such as Latin and (ii) under-represented languages. Second, while we have seen some psycholinguistic evidence that probabilistic grammars are deep and real (again, only with English data on a well-known alternation), there is to my knowledge no neurolinguistic experimentation of probabilistic models. For instance, to what extent is the most probable variant (given the discourse and sentential context) faster to access, and is there a focus positivity effect found with the less probable variant? I think that this might be worth investigation. Third, as discussed by de Marneffe et al. (2012), there is strong evidence that children pick up probabilistic cues, but how long does it take for the adult probabilistic grammar to fully emerge. Fourth, to what extent do L2 speakers identify the same constraints as L1 speakers? Furthermore, for a language like Latin, where there are no native speakers left, are L2 speakers of Latin sensitive to any of the constraints that Latin native speakers were sensitive to, as identified by corpus analysis? Finally, the focus has been on alternations. However, to what extent are other linguistic phenomena (such as the identification of grammatical distinctions between subject and object and informational distinctions between focus and background and between topic and comment) sensitive to probabilistic and multivariate information as discussed in this chapter?

With this discussion in place, we turn in the next part of the thesis to a detailed empirical case study, addressing the applicability of probabilistic and multivariate models to classical Latin data.
Part II

Data and Materials
Chapter 4

Defining and Delimiting the PAC for Operational Purposes

4.1 Introduction

In this part of the thesis, we turn to matters related to the empirical fundamentals of the present study, viz. the collection of data and materials.

We have thus far only talked of the PAC in very general terms, without defining it. We begin with the present chapter that gives an operational definition of the PAC to guide the derivation of the datapoints. I shall first discuss empirical problems with how previous investigators of the construction under analysis have attempted to define what I term PAC, and then delineate what it constitutes for the purpose of the present investigation. As will be emphasized this is crucial for a replicable study.

4.2 Previous definitions and problems

Previous investigators of grammatical variation within the construction that concerns the present study, i.e. what I have been calling thus far the “participle-auxiliary construction (PAC)”, have either not stated, or stated in vague terms, what they understand by the terms “participle” and “auxiliary” and the construction they are part of. Consider the following two cases.

Devine & Stephens (2006) nowhere define what exactly they mean by the terms “participle” and “auxiliary”, and more precisely they do not state what types of participle are included in the remit of their investigation. An examination of their data, however, reveals that they include mainly perfect participle constructions, e.g. dimissa est ‘was dismissed’. A small number of future participles are also included (five at (2006: 183, ex. (87))), but only when they are in infinitival forms (of the kind facturum esse ‘to be going to do’) and in negative contexts and with esse preceding the participle. Other possible participle types, such as the gerundive, are not mentioned in the section on the location of the auxiliary. Thus, it is unclear what the structure under consideration comprises.

Spevak (2010), by contrast, does state the object of her investigation with references to the structures’ inclusion in a verb’s paradigm: “[t]he passive forms of the perfect and the pluperfect tenses are analytic and consist of a passive past participle and the auxiliary sum” (2010: 149) and “[d]eponent verbs in the perfect and pluperfect tenses show morphological affinities with the passive voice. Therefore, I will examine them from the same point of view” (2010: 155). In other words she is going to go through a corpus and extract structures like uictus est ‘he has been beaten (by x), he was beaten (by x)’ and uictus erat ‘he had been beaten (by x)’ which are the passive correspondences of the active forms uicit ‘he has beaten x, he beat x’ and uicerat ‘he had beaten x’, respectively, as well as things like mortuus est ‘he has died, he died’ and mortuus erat ‘he had died’, which is the only possible way of expressing the
perfect and pluperfect tenses in deponents.\textsuperscript{1} While this might at first sight seem sufficient intuitively, it turns out to be problematic on an operational level. When Spevak collected her data she encountered the data point \textit{L. Domitius ... ab equites est interfectus} ‘Lucius Domitius was killed by the cavalry’ (Caes. Civ. 3. 99; 5; Spevak (2010: 154, ex. 22)). She read the context, and positively identified it as being an “analytic” perfect tense, and included it in her dataset: the PAC \textit{est interfectus} means ‘was killed’ and is the passive form for the perfect tense of \textit{interficere} corresponding to the active form, e.g. \textit{L. Domitium equites interfecerunt} ‘the cavalry killed Lucius Domitius’. So this is a clear-cut case, as are all of the other 6 examples for the passive and all 2 examples for the deponent reported in the actual discussion.\textsuperscript{2} However, given that she reports only 3.4\% (7/150+55) of her dataset for the passive and 2.2\% (2/70+19) of her dataset for the deponents, it far from clear how she dealt with the other 96.6\% and 97.8\%, respectively, particularly in those cases where the decision as to whether a particular structure should be included in the dataset or not is not as straightforward. Consider now the case of (1), which to judge from her brief corpus list should have been encountered during her collection of the data.

\begin{center}
\begin{tabular}{ll}
\hline
(1) & adeo\textsuperscript{so} et \textit{impedita} \textit{ut} ... \\
& so\textsuperscript{so} and \textit{be.3SG.IMPF.IND obstruct.PTCPL.PRF.F.SG.NOM valley.F.SG.NOM that} ... \\
& ‘And the valley had been obstructed so that ...’

(Caes. Civ. 2. 34. 1)
\end{tabular}
\end{center}

Here \textit{erat impedita} can be interpreted as being the passive pluperfect tense of the verb \textit{impedio} corresponding to the active pluperfect tense form \textit{impediuerant}, and would therefore fit the definition given above. However, in the context there is no expression of agency to mean that ‘\textit{x} obstructed the valley’; the agent, if there ever was one, goes unspecified. Accordingly, one can also interpret this sentence as ‘and the valley was so obstructed that’, and this is exactly how it is taken by Devine & Stephens (2006: 182; see their translation). On this interpretation, then, \textit{impedita} is effectively an adjective + copula-\textit{esse} construction (for \textit{impeditus} as an adjective, cf. \textit{OLD} (Glar, 1982) s.v. \textit{impeditus}), and not an analytic form that is part of the paradigm of \textit{impedio}.\textsuperscript{3} The question that arises is this: assuming that Spevak did encounter this example, did she include it because she thought it was pluperfect, as it can be on one interpretation, or did she exclude as being an imperfect copula + adjective structure, as it can be on the other interpretation, or did she vacillate on the issue? The only person that can answer this question is Spevak herself, as she does not provide for those who want to re-examine her data a definition of inclusion that is fit for purpose: it only works for clear-cut cases, and is totally useless when ambiguous cases arise.\textsuperscript{4}

This lack of clarity with regard to the delineation of the construction flies in the face of what is required for a systematic and replicable investigation, and crucially makes it difficult for these scholars’ hypotheses to be tested quantitatively.

The take-home point here is that it will be necessary in the present investigation: (i) to give an explicit, operational definition of what is understood by the term PAC which takes into account the possibility of ambiguity between verbal and adjectival participles, and crucially one that does not involve reference to the paradigm; and (ii) to list a set of participle + \textit{esse} sequences that can conceivably considered a PAC, based on a set of operationally more objective criteria, and give reasons for inclusion and exclusion.\textsuperscript{5}

\textsuperscript{1}I.e. passive forms with active meanings.

\textsuperscript{2}Not including instances of \textit{factus} \ldots \textit{est} or \textit{est} \ldots \textit{factus}.

\textsuperscript{3}The distinction between verbal (eventive) and adjectival (stative/resultative) participles will be addressed in more detail in a subsequent chapter.

\textsuperscript{4}\textit{est autem oppidum et loci natura et collie munimentum} “However, the town was protected by the nature of its site and by a hill” (Caes. Civ. 3. 9. 2). Although this is an \textit{est}...\textit{factus} variant, i.e. not one investigated here, it is interesting as Spevak includes it in her comments on adjectives + \textit{esse}, but goes and translates it as if it were an analytic perfect passive PAC.

\textsuperscript{5}It is essential to contrast the difference between an operational definition – one that can be used in a quantitative analysis – and a theoretical definition. For instance, what I consider a participle here may not be what theoretical morphologists and syntacticians consider a participle. For instance, I consider \textit{certus} (‘certain’ and \textit{ratus} (in the meaning ‘fixed’) participles, simply because of their historical form. No attempt will be made to define a PAC theoretically in this investigation.
4.3 Definition in the present study

In this investigation, I define a participle-auxiliary construction (PAC) operationally as the following: a bipartite verbal construction comprising a “participle” and an “auxiliary”, syntactically continuous or discontinuous, where the term “participle” is to be understood as a deverbal adjectival form that inflects via agreement for the case of the subject (i.e. nominative or accusative, the case forms in which it can occur with esse), for the gender of the subject (i.e. masculine, feminine, and neuter) and for number of the subject (i.e. singular and plural), and where the term “auxiliary” is simply the label for a form of esse which co-occurs with a participle so understood.

This loose definition has some undesirable consequences, though. Since this is not a definition based on a verb’s paradigm, which I argued to be problematic above, it includes not only the perfect participle, the present participle, the future participle, and the gerundive but also various formations in -bundus, -cundus, -dis and the extended form -bilis, -nus, -mus, -az, -icius and -idus. In order to narrow down the definition, I specify a condition on the PAC that it includes those deverbal forms that: (i) are frequently and well-attested with esse, and (ii) have a wide lexical distribution when in combination with esse. As we will see, this reduces the definition to include only the extensions of (i) the perfect participle, (ii) the future participle, and (iii) the gerundive.

I now go through each of the three participles mentioned, stating why they are included; the criteria by which they are selected, and mention some problematic cases. After this, I briefly state why those other forms are not going to be included, even though they fit my originally loose definition.

4.3.1 Perfect participle + esse

The first participle to be included is the perfect participle in -tus (including the morphophonological variant -sus and the form -tuus, which is present in the participle mortuus ‘dead’, for instance), regardless of whether it is used in a clear-cut adjectival sense (with esse functioning as a copula), a clear-cut verbal sense (with esse functioning as auxiliary), or in an ambiguous adjectival/verbal sense (with esse thus having an ambiguous syntactic function). There are several reasons why its inclusion is warranted. First, the combination of perfect participle + esse is well attested from the beginning of the Latin record, and it has parallel forms in other Indo-European languages, including instances in Italic. Second, perfect participles that enter into collocations with esse have a wide lexical distribution. Let us now examine some criteria by which perfect participle forms were selected.

1. First, as will be obvious from my definition of a PAC, the participle form in -tus, -sus, or -tuus must be derived from a verb. The -tus suffix is a productive one throughout the history of Latin, not only as an adjectivalizing/participializing suffix built on to verb stems but also as an adjectivalizing one built on to nominal stems. In this investigation, I include the former but not the latter, except in cases where it is unclear whether the formation derives from a nominal or a verb.

In most cases, it is clear that we are dealing with the verb-derived -tus form. For instance, amatus est ‘he is/was loved’ is derived from a verbal stem in ama-, and is therefore included. The participle interfectus ‘killed’ is derived from interficio ‘kill’, so this example is included. The participle factus ‘done, made’ is derived from facio ‘do, make’, so again this example is included.

A case like paratus ‘prepared’ is also included, as it is derived from the verbal stem para-.

In other cases this is not so. These include formations in -tus which are derived from a nominal base, such as the notional barbatus est ‘he is with-a-beard’ (← barba ‘beard’ + -tus, -tus is ‘endowed-with’), and clipeatus est ‘he is with-a-shield’ (← clipeus ‘shield’ + -tus (e.g.) ‘endowed-with’). Such

---

6 Though this is not relevant here, as I am only looking at continuous constructions.

7 See the discussion later on this distinction.

8 See Buck (1933), where forms such as präftüset (Oscan) and kuratu eru (Umbrian) are mentioned.

9 Factus is also the perfect participle for fio ‘I become, am made’.
cases are thus excluded from the investigation’s portfolio.\footnote{The literature on Dutch participles reports that “pseudo-participles”, which are formally identical to participles but are related to no attested verb stem/infinitive, must be considered adjectives. For example, Cousse (2011: 616) points out that gespierd ‘muscular, provided with muscle’, while formally identical to a participle (note the ge- inseparable prefix), does not derive from a verbal stem, but rather spier ‘muscle’, which is a noun. Such instances are included in the dataset of De Sutter (2005, 2009).} Additionally, there are forms which waver between a noun-derived form and a verb-derived form. Is the structure armatus est, for example, derived directly from nominal arma ‘arms’ + -tus ‘endowed-with’ or from a verb stem in arma- (cf. armare) with a participializing -tus? It is impossible to tell. Thus, since the participle can go back to an attested verb, such forms are included.\footnote{An interesting case is provided by operatus, which might be cast into doubt for etymological reasons. Housman (1972: 1235) alleges that this form in the context in which it is described (Prop. 2. 32. 2) is not a perfect participle, but rather a nominally-derived adjective built on opera- (noun) + -tus (adjectivalizing). The reason for this is that operatus is attested much earlier than the first attestation of the verb operor ‘I engage in work’ from which a participle form operatus ‘engaged in work’ might be derived. Since the verb is actually attested (unlike *barbo or *clipeo), it is less certain that the form reflects a nominal rather than a verbal derivation. We cannot be certain that the verb was not in usage before Propertius.}

2. Also included are perfect participle forms that are historically associated with a verb’s lemma, but are semantically and/or morphophonologically distinct in some way. Here, we find, for instance, altus ‘having great extension upwards, lofty, tall, high’, contentus ‘content, satisfied’, privatus ‘restricted for the use of a particular person, private’, situs ‘laid up, stored, deposited’. These participles are historically related to their verbs, alo ‘suckle, nurse, feed’, contineo ‘hold together, connect, link, join’, priuo ‘cause to be parted (from), deprive or rob (of)’, sino ‘leave alone, let be’, respectively, but — for whatever historical reason — their meanings have become metaphorical or specialized.

3. I mention a final point here concerning the inclusion/exclusion of other -tus, and morphophonologically related, forms. A number of perfect participles + esse, such as paratus est, have (i) forms derived from themselves which show negative prefixation with in- (e.g. imparatus est ‘he is un-prepared’) and (ii) inflectional comparative and superlative forms (e.g. paratior est ‘he is more prepared’, paratissimus est ‘he is most/very prepared’) besides the basic positive. Since paratus est can go back to a listed verb parare ‘to prepare’, this will be included in the investigation’s data portfolio. But since there is no listed corresponding verb *imparare attested with the meaning ‘to make unprepared’ or *paratiorare ‘to make more prepared’ or *paratissimare ‘to make most prepared’ forms such as imparatus est, paratior est, paratissimus est will thus be excluded from the investigation, as to include them would be in direct conflict with my stipulation in 1, above.

4.3.2 Future participle + esse

The future participle encoded morphologically by -turus, -surus, etc. will be included for the following reasons. The future participle has restricted independent usage outside the participle + esse structure, whether in attributive or predicative functions, and these functions are attested much later than its function in connection with esse.\footnote{For a recent discussion of this collocation, see Vincent (2011).}

The selection of future participles is essentially unproblematic as there are very few types ending in -turus which are derived from forms other than an attested verb. Thus, moriturus ‘going-to-die’ is associated with the lemma morior ‘to die’, and the construction moriturus est ‘he is going to die’ is included.

4.3.3 Gerundive + esse

The third type of participle to be included in the dataset is the gerundive — this participle indicates obligation and necessity\footnote{When it is used predicatively with or without esse.} and is morphologically marked by -ndus (and morphophonological variants).
4.3. DEFINITION IN THE PRESENT STUDY

One strong argument for a gerundival type of *pacc* is the fact that the gerundive functions differently when combined with *esse* (e.g. in subject function) from when it occurs in other syntactic environments. This is well in evidence in both personal structures, as in for instance *frater est exspectandus mihi* ‘I should wait for my brother’ (Ter. Ph. 460) and in impersonal structures, as in *apparebat aut hostibus aut ciuibus de victoria concedendum esse* ‘It was apparent that either the enemy or the citizens had to concede victory’ (Liv. 4. 6. 6). By contrast, in other functions, there is no such modality nuance, and the gerundive is simply a passive verbal adjective, as in *dolum ad capiendos eos comparant* ‘they prepare a trap for them to be taken’ (Liv. 23. 35. 1) and *quom diuiti homini id aurum seruandum dedit* ‘when he gave this gold to a rich man to be looked after’ (Pl. Bac. 337). Like both the perfect and future participles + *esse*, the gerundive + *esse* structure is attested widely throughout the lexicon, and not restricted to just a handful of lexemes. Let us now proceed to a more detailed itemization of what was included and what was not.

1. As with perfect and future participles, for the gerundive form to be included in this investigation, it must be derived from or formed on an attested verb.

2. Gerundives are accepted only if they combine with *esse* and result in a necessity/obligation interpretation. This includes those gerundives which, because they are often interpreted in a specialized way and/or metaphorically, have a separate lemma in *OLD* (Glare, 1982) — e.g. *abominandus* ‘to-be-averted-by-omen, of-evil-omen’, *admirandus* ‘to-be-surprised-at, astonishing’, *conspiciendus* ‘to-be-caught-sight-of, conspicuous’. Although these have a separate lemma, they still retain some degree of modal force in combination with *esse*. By contrast, other gerundives, which are also listed separately from the verb lemma in *OLD*, cannot bear such an interpretation — e.g. so-called old gerundives from deponent verbs such as *labundus* ‘gliding’, *oriundus* ‘originating’ and *secundus* ‘following, favourable’, and possibly *rotundus* ‘rounded’. These forms will not be included for the following reasons. When combined with *esse*, they do not have a specialized modal function like regular gerundives + *esse*. For instance, *oriundus est* is used as an alternative for the one-word form *oritur*, as in e.g. *inde sum oriundus* ‘I originate from there’ (Pl. Poen. 1053) vs. *Miseno oriuntur echini* ‘sea-urchins originate from Misenum’ (Hor. S. 2. 4. 30). But the structure has only present reference; there is no modality of necessity/obligation connected with it, even though the form is combined with *esse*. Furthermore, they do not have a widespread functional distribution throughout the verbal lexicon and are restricted to a handful of lexemes.

3. There are gerundive forms with the negative prefix *in- ‘un-‘ — e.g. *impaentiendus* ‘not to be repented of’, *incredandus* ‘not to be believed’, *infandus* ‘unspeakable’, *intemerandus* ‘inviolable’, and *intolerandus* ‘insupportable’. These forms + *esse* will not be included for the same reasons as given above for *imparatus* ‘unprepared’.

4.3.4 Exclusion of other deverbal forms + *esse*

As we saw from the definition above, next to the three participle types mentioned, Latin also has a number of other deverbal forms which might potentially also occur in the environment of *esse* ‘to be’ — e.g. present participles in *-ns*, verbal adjectives in *-bundus, -cundus, -lis* and the extended form *-bilis, -nus, -mus, -az, -icius* and *-idus*. The definition did not exclude these forms, but I will not include them on distributional grounds.

Examples (2)–(3) are instances of present participles in *-ns + esse*. In (2) *audiens* ‘hearing’ is morphologically derived from the verb *audio* ‘to hear’; in this example, *audiens* participates in the idiomatic collocation *dicto audiens* ‘obedient’. The type *sciens est* in (3) involves the present participle of *scio* ‘to know (a fact)’.
The combination of present participles + esse has a tendency to appear in recurring lexemes. I list some from Plautus and Terence: obsequens est (e.g. Pl. Curc. 258, Pl. Merc. 150, Pl. Merc. 158); oboediens est (e.g. Pl. Mil. 806); (dicto) audiens est (e.g. Pl. Pers. 399, Pl. Per. 836, Pl. Trin. 1061); sciens est (e.g. Pl. Poen. 1038, Ter. An. 508, Ter. An. 775). Some of these structures are restricted to non-core meanings of the verb form from which it is derived: audiens est means ‘he is obedient’, but it is not used in the basic sense of the verb ‘he is hearing’ (OLD s.v. audio).

Examples (4)–(5) are generally held to be deverbal adjectives, so if they are found with esse they would meet the broad definition of a PAC given above. The form calidum in (4) ‘hot’ is a deverbal of the intransitive verb caléo ‘to be heated’ with an -idus suffix. In (5), terribilis is a deverbal of the intransitive terreō ‘to terrify’ formed from a base and the suffix -bilis.

Data searches show that collocations of these verbal adjectives + esse are relatively rare against their usage in other syntactic functions — e.g. in attributive or predicative contexts. There are also some distributional restrictions on their forms: -idus est occurs only with a finite range of intransitives, including auidus est ‘he is eager’ (cf. auo ‘to be eager for’), lucidus est ‘he is light’ (cf. luco ‘to be light’), and timidus est ‘he is timid’ (cf. timeo ‘to be fearful’).

### 4.4 Final remarks

In this chapter we have set out the types of participle and esse ‘to be’ sequences that will feature in the present study. They include perfect participles, whether of a verbal or adjectival nature, future participles, and gerundives. Other deverbal formations, such as the present participle, and deverbal suffixes of the type -bilis, -mus, -mus, -ax, -icius are excluded from the analysis, because they do not frequently occur with esse, and there would be limited occurrences of them in the dataset to probe statistically. These choices may strike the reader as being arbitrary, to some extent, but it is necessary to delimit the response structures of interest, to narrow the scope and allow replicability. With this in place, we turn in the next chapter to the choice of data.
Chapter 5

The Choice of Data

5.1 Introduction

The choice of data in the present study is Latin prose from the Late Republican period (90–27 BC), passed down to us in the literary record. This choice is motivated positively by three main reasons. First, this period is typically associated with “Classical Latin”, a type we are most familiar with from school, and thus the type most accessible to non-specialists of Latin. Second, it has been argued that in the Late Republic the est factus serialization gains the greatest ground on the factus est serialization (see e.g. Marouzeau 1910). And third, it has been reported that there was a “progressive evolution of the passive voice into a fully-fledged paradigmatic alternative to the active” in this period (Clackson & Horrocks, 2007: 213), a voice which uses constructions with esse ‘to be’ in perfect forms — we might therefore expect to find a high textual concentration of PACS within this period, thus facilitating the collection of as many relevant data-points as possible.

Before I state which authors and work comprise the present dataset (5.3), I shall first provide reasons (i) why inscriptions (5.2.1) and poetry (5.2.2), other potential loci of the variability in our construction, and (ii) why categorical authors (5.2.3) are excluded.

5.2 Excluded datasources

5.2.1 Inscriptions

In addition to the literary record, there is a good deal of inscriptional evidence for word order alternation in PACS. For example, in CIL I.2 756. 11–12, the factus est and the est factus variants occur in close succession in separate subordinated clauses that revolve around an imperatival matrix clause: quae pequnia ad eas res data erit, profana esto, quod d(olo) m(alo) non erit factum, ‘which money will have been given for those things, let it be profane if it will not have been done with malicious intent’. In spite of inscriptions being a potentially expansive data source, we must nonetheless exclude it. First, unlike literary data, they are not richly available in digital format at present; this would unfortunately necessitate hours of manually reading extraneous material to source observations. Second, many of the inscriptions are often only snippets without context, with words missing. As such the observations would be difficult to annotate for a number of contextually important variables, such as persistence and information structure.

---

1See Kruschwitz (2004) for a good discussion (with reference to Wackernagel’s law). Besides instances of the continuous factus est and est factus variants, cited here, Kruschwitz also discusses discontinuous variants.
5.2.2 Poetry

Alongside its incidence in prose texts, word order variability in *pacs* is in evidence in the work of poets of the late republic. Poetry will not be included here, though, as a pilot study indicated that in 71 observations sampled from *De Rerum Natura* of Titus Lucretius Carus (c. 98–55 BC), metrical constraints were found to be pervasive, forcing the word order of 92.96% of the observations.

We will see later that certain prose texts (e.g. that of Cicero) may be “metrical” in the sense that certain period-closing patterns of heavy and light syllables are preferred (cf. e.g. Clackson & Horrocks 2007: 221ff.). Thus, the order in (1a) — the textually attested variant — may be preferred because it results in a “more preferable” metrical pattern (cretic + trochee), whereas (1b) — the one that I reversify — results in a less preferred metrical pattern (spondee + trochee).

(1) a. respondit a Romilia tribu se initi(um) essē fāctūrum.
   b. respondit a Romilia tribu se initium fāctūr(um) essē.

The principles in operation in prose are much more difficult to set down as clearly as they are in verse, and it is consequently much more subjective to analyse them. With this caveat in mind, I admit the evidence of only prose in this investigation.

5.2.3 Categorical authors

Only individuals that exhibit variability between *factus est* and *est factus* are included in the present dataset. An important writer of Late Latin Republican prose, the historian Sallust (86–35/34 BC) should be mentioned in this regard. Vogel (1937: 64), Devine & Stephens (2006: 216ff.), Brookes (2009: 66), and Spevak (2010: 151) observe that the *est factus* variant — alongside fronted instances of *esse* more generally — is vanishingly rare in this author. According to data in Brookes (2009: 65), Sallust exhibits 254 instances of *factus est* versus only 3 instances of *est factus*. These figures correspond to a proportional rate of 98.84% and 1.16%, respectively, thus indicating that Sallust is virtually a categorical individual in his choice of variant. For this reason, it is obvious that his works must be excluded from the present dataset.

5.3 Included datasources

This section introduces the authors and works that will be used to source the data-points.

5.3.1 Cicero

Marcus Tullius Cicero (106–43 BC) provides the most extensive surviving corpus of any Roman author to have been transmitted to us. Besides his oratorical works (both political and forensic), he wrote on rhetoric, philosophy, and politics, and was an energetic correspondent. For the purposes of this investigation, I focus on a central genre — oratory (*orationes*).

Observations are sourced from the *Philippics*, a set of 14 deliberative political speeches against Marcus Antonius, which were delivered between 44 and 43 BC. The extant *Philippics* have a textual span of 53396 words (according to *LLT-A*, see Chapter 6), and are the “only thematically connected and the largest coherent corpus of Ciceronian political speeches extant” (Manuwald, 2008: 65).

Let us now turn to some general features of Cicero’s language. With a deep understanding of grammar and rhetoric, Cicero was a pre-eminent stylist, who is generally associated with the standardization of the Latin language. In this regard, Cuzzolin & Haverling (2009: 46) observe “In Late Antiquity Cicero...”

---

2 There were probably 18 original speeches, cf. Manuwald (2008).
3 See Manuwald (2008) for the sociohistorical context.
is often referred to as the great unsurpassed model of Latin prose and oratory and eventually his Latin — along with that of, e.g. Caesar — was “virtually the only standard by which all Latin texts had to be judged with regard to correctness”. Thus this “correctness”, or *Latinitas* as it is sometimes known, is a key linguistic feature of Cicero’s works, generally envisaged.

### 5.3.2 Caesar

Gaius Iulius Caesar (100–44BC) wrote a number of works, including speeches, grammatical treatises, and verse compositions. However, he is most famous for his two extant commentaries, the *Bellum Gallicum* ‘Gallic War’ and *Bellum Civile* ‘Civil War’. These two works will be the focus of the present investigation. The former work deals with Caesar’s subjugation of Gaul, comprises 7 books, and has a textual span of 45790 words. The latter deals with the civil war that was started in 49BC, comprises 3 books, and has a textual span of 33268 words.

Caesar has been discussed in respect of the *pac* alternation, in e.g. Vogel (1937), Devine & Stephens (2006) and Spevak (2010).

### 5.3.3 Caesar’s continuators

Several writers continued Caesar’s commentaries, and are thus known as his “continuators”. I include their works in the present dataset. These include Aulus Hirtius (consul 43BC), an officer of Caesar, who added an eighth book to *Bellum Gallicum* to provide a thematic transition between this work and *Bellum Civile*. This work has a textual span of 6639 words. Three works continue the *Bellum Civile*, and are of uncertain authorship. They are, in order of their position on the manuscript, *Bellum Alexandrinum*, which comprises 10666 words, *Bellum Africum* with a textual span of 13484 words, and *Bellum Hispaniense* with a textual span of 6168 words.

In terms of language, they differ strikingly from Caesar’s commentaries, especially the *Bellum Hispaniense*, which, written presumably by a soldier, “is one of the very few works written in a predominantly un-literary Latin, and is, therefore a very valuable source for knowledge of the language” (Kenney & Clausen, 1982: 285).

### 5.3.4 Nepos

Cornelius Nepos’ dates cannot be determined with accuracy, but best estimates suggest a date of birth of c.100 BC. Nepos’ literary output was reportedly vast, ranging from chronologies, collections of anecdotes, geographical treatises, and biographies. For the purpose of this investigation, I draw on the surviving parts of Nepos’ *De Viris Illustribus* (‘On Famous Men’). More specifically, *pac* tokens are taken from the extant book on foreign military leaders (*De Excellentibus Ducibus Exterarum Gentium* containing 23 lives, and two lives from *De Historicis Latinis* (‘On Latin Historians’). These works have a textual span of 28811 words in all. I note that Nepos’ works have not yet been scrutinized with respect to the *pac* alternation.

### 5.3.5 Varro

Marcus Terentius Varro (116–27 BC), the final author under consideration, wrote prolifically. Of the two works of his that survive, the present dataset is drawn from *De Re Rustica*, a treatise on agriculture written in the form of fictitious conversations, which include dramatic interruptions. This work is partitioned into three books. Book 1 deals with general aspects of agriculture, and is dedicated to his wife. Book 2 concerns the treatment of livestock, and is dedicated to a rearer of cattle. Book 3, which is dedicated to a neighbour, turns to the remaining topics of farming — apiculture and fish, for instance. In total, this work has a textual span of 35692 words. There are some brief remarks on Varro’s *pac* usage in Kroll (1921), but, to my knowledge, there is as yet no widescale, quantitative investigation.
5.4 Conclusion

This chapter has motivated the presence of nine prose texts of Late Republican Latin for inclusion in the present study: the anonymous works of Bellum Africum (B. Afr.), Bellum Alexandrinum (B. Alex.), and Bellum Hispaniense (B. Hisp.); Caesar’s Bellum Civile (Civ.) and Bellum Gallicum (Gal.); Cicero’s Phillipics (Phil.); Nepos’ De Viris Illustribus; and Varro’s De Re Rustica (RR). Table 5.1 lists the text sources used in this investigation and includes word counts for the respective sources. Due to certain restrictions, epigraphic and poetic data cannot be taken into consideration.

<table>
<thead>
<tr>
<th>Author/Work</th>
<th>Word Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymous B. Afr.</td>
<td>13484</td>
</tr>
<tr>
<td>Anonymous B. Alex.</td>
<td>10666</td>
</tr>
<tr>
<td>Anonymous B. Hisp.</td>
<td>6168</td>
</tr>
<tr>
<td>Caesar Civ.</td>
<td>33268</td>
</tr>
<tr>
<td>Caesar Gal.</td>
<td>45790</td>
</tr>
<tr>
<td>Cicero Phil.</td>
<td>53396</td>
</tr>
<tr>
<td>Hirtius Gal.</td>
<td>6639</td>
</tr>
<tr>
<td>Nepos Vitae</td>
<td>28811</td>
</tr>
<tr>
<td>Varro RR</td>
<td>35692</td>
</tr>
<tr>
<td><strong>Total Word Count</strong></td>
<td><strong>233914</strong></td>
</tr>
</tbody>
</table>

Table 5.1: Text sources and word counts
Chapter 6

Data Extraction

6.1 Electronic text archive

I elected to use *Library of Latin Texts — Series A* (henceforth: *LLT-A*) (Tombeur, 2012). The *LLT-A* is an unannotated, lemmatized, machine-searchable text archive, which was established as *Cetedoc Library of Christian Latin Texts* (*CLCLT*) in 1970, and at present comprises 69.17 million word-forms spanning 3394 works (Tombeur, 2012: 6). It is stratified according to period, century, author, and title of work (Tombeur, 2012: 9). The text on which this archive is based comes from published classical editions (Tombeur, 2012: 40ff.); the ones pertinent to this investigation are reported in table 6.1 below.

<table>
<thead>
<tr>
<th>Author/Work</th>
<th>Edition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymous B. Afr.</td>
<td>Klotz (1927)</td>
</tr>
<tr>
<td>Anonymous B. Alex.</td>
<td>Klotz (1927)</td>
</tr>
<tr>
<td>Anonymous B. Hisp.</td>
<td>Klotz (1927)</td>
</tr>
<tr>
<td>Caesar Civ.</td>
<td>Klotz (1950)</td>
</tr>
<tr>
<td>Cicero Phil.</td>
<td>Fedeli (1986)</td>
</tr>
<tr>
<td>Varro RR</td>
<td>Goetz (1929)</td>
</tr>
</tbody>
</table>

Table 6.1: Textual editions used in *LLT-A* for the texts considered in this study

6.2 Identification of data points

To obtain the required information, a number of filtration stages were employed. In the first step, I used the *LLT-A* platform to establish how many potentially relevant token sites there are. This is equivalent to the number of times orthographic forms associated with the verb lemma *esse* ‘to be’ occur in the texts we are interested in.

The result of this search reveals that, for Caesar’s *Bellum Gallicum*, there are 762 contexts or *sententiae* in which at least one orthographic form of the verb-set *esse* is used (N.B. a *sententia* is a chunk of text which is more or less equivalent to “sentence”; see Tombeur 2012). I then went through these *sententiae*, and manually extracted all those forms which had as an immediate left or an immediate right collocate one of the participles mentioned above in Section 4.3, i.e. perfect participles, future participles, and gerundives, in accordance with the following schemata.

---

1 Of course other choices might have been made. For instance, Spevak (2010) and others have made use of LASLA at [http://www.cipl.ulg.ac.be/Lasla](http://www.cipl.ulg.ac.be/Lasla). This would be problematic for the present investigation, as it does not contain the works of Nepos or of Varro.

2 I use “text archive” in the sense of Gries (2009), who distinguishes it from “corpus” in the technical sense.
CHAPTER 6. DATA EXTRACTION

(1) a. [participle] esse ← factus est
   b. esse [participle] ← est factus

This process deletes cases such as the following, where esse does not have a participial left or right collocate. In (2) sunt, which has identificational semantics, is flanked by a pronoun hi and a preposition extra, respectively, whereas in (3) est, which has existential semantics, is flanked by two nouns.

(2) hi sunt extra provinciam trans Rhodanum primi.
   ‘These are the first [peoples] beyond the province across the Rhone.’ (Caes. Gal. 1. 10. 5)

(3) flumen est Arar
   ‘There is a river, Arar...’ (Caes. Gal. 1. 12. 1)

Serializations with any degree of discontinuity, such as factus... est and est... factus, were accordingly deleted, because they are not within the scope of the project, as has been mentioned earlier.

In some cases, there is a participle flanking the form of esse on the left and on the right. Thus at first sight, it is unclear whether we are dealing with factus est or est factus.

(4) idem=que temporis glans missa
    same.N.SG.ACC=and time.N.SG.GEN missile.F.SG.NOM send.PTCPL.PRF.F.SG.NOM
    est inscripta
    be.3SG.PRS.IND inscribe.PTCPL.PRF.F.SG.NOM
    A: ‘a missile was sent, inscribed...’
    B: ‘a missile, having been sent, was inscribed...’ (B. Hisp. 13. 3)

Here we might have a PAC with the factus est variant in missa est, reading A, or a PAC in the est factus variant with est inscripta, reading B. In this case, the former interpretation is more appropriate, mainly because the narrative concerns the temporal sequence of events (the expression idemque temporis ‘and at that moment in time’ signals that the narrative is to be pushed forward, and inscripta simply provides backgrounded information). Such cases were determined contextually on a case-by-case basis.

6.3 Textual issues

During the process of uploading texts into the LLT-A database, errors inevitably occur, a point mentioned in the user’s manual (Tombeur, 2012: 18). Second, the online edition offers no critical apparatus, so it is difficult to determine whether there are any problematic textual points that obviously need to be raised. I therefore compared each of the extracted PACs with its co-referent in a number of other editions, where possible, to check for editorial decisions when a textual choice arose as a result of transmission. Table 6.2 indicates the editions that were used for comparative purposes, and Appendix A illustrates a number of cases. How should we deal with textual issues? With the caveat in mind that all texts that are transmitted to us from antiquity involve some textual anomalies, I will include these data points in the analysis, since they are rather trifling and affect only a small proportion of the dataset.

6.4 Frequency overview

The dataset of the PAC alternation encompasses in all \( n = 2409 \) data-points, of which 1450 (60.19%) are in the factus est variant and 959 (39.81%) are in the est factus variant. The difference between these proportions is highly significant (\( \chi^2 = 100.08, d.f. = 1, p < 0.001 \)), a finding that is in line with previous scholarship, where the factus est variant has been reported to be the more frequent, as in e.g. Hoff (1996: 374), Brookes (2009: 10, 65–66) and Spevak (2010: 151, 155). Notably, the factus est form,
<table>
<thead>
<tr>
<th>Author/Work</th>
<th>Edition</th>
<th>Editions used for comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymous <em>B. Alex.</em></td>
<td>Klotz (1927)</td>
<td>Du Pontet (1900), Andrieu (1954)</td>
</tr>
<tr>
<td>Varro <em>RR</em></td>
<td>Goetz (1929)</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 6.2: Editions used for textual comparison
Table 6.3: The overall distribution of *factus est* and *est factus* variants cross-classified by the authors/works under investigation.

<table>
<thead>
<tr>
<th>Author/Work</th>
<th>Frequency (and %) of <em>factus est</em></th>
<th>Frequency (and %) of <em>est factus</em></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymous B. Afr.</td>
<td>45 (42.06%)</td>
<td>62 (57.94%)</td>
<td>107 (100%)</td>
</tr>
<tr>
<td>Anonymous B. Alex.</td>
<td>62 (56.36%)</td>
<td>48 (43.64%)</td>
<td>110 (100%)</td>
</tr>
<tr>
<td>Anonymous B. Hisp.</td>
<td>57 (44.88%)</td>
<td>70 (55.12%)</td>
<td>127 (100%)</td>
</tr>
<tr>
<td>Caesar Civ.</td>
<td>125 (48.64%)</td>
<td>115 (41.36%)</td>
<td>257 (100%)</td>
</tr>
<tr>
<td>Caesar Gal.</td>
<td>214 (65.05%)</td>
<td>115 (34.95%)</td>
<td>329 (100%)</td>
</tr>
<tr>
<td>Cicero Phil.</td>
<td>480 (74.65%)</td>
<td>163 (25.35%)</td>
<td>643 (100%)</td>
</tr>
<tr>
<td>Hirtius Gal.</td>
<td>33 (48.53%)</td>
<td>35 (51.47%)</td>
<td>68 (100%)</td>
</tr>
<tr>
<td>Nepos Vitae</td>
<td>326 (60.26%)</td>
<td>215 (39.74%)</td>
<td>541 (100%)</td>
</tr>
<tr>
<td>Varro RR</td>
<td>108 (47.58%)</td>
<td>119 (52.42%)</td>
<td>227 (100%)</td>
</tr>
<tr>
<td>Total</td>
<td>1450 (60.19%)</td>
<td>959 (39.81%)</td>
<td>2409 (100%)</td>
</tr>
</tbody>
</table>

Figure 6.1: Descending proportions of *factus est* by authors/works

described as the “banal order” by Adams (1994: 40) and “basic” by Brookes (2009: 66), is the one found in the discussions of ancient grammarians on verbal constructions, and it is the form we learn from the paradigms in the standard school grammars, e.g. in Kennedy & Mountford (1962: 61, 72–73).

Table 6.3 displays the distribution of *factus est* and *est factus* variants cross-tabulated by the texts in alphabetical order. The first line of table 6.3, for instance, reports the frequencies for the anonymous author of the *B. Afr.*, indicating that there are 45 instances of *factus est* (corresponding to a proportional rate of 42.06%) and 62 instances of *est factus* (corresponding to a proportional rate of 57.94%). This cross-tabulation is displayed graphically in the barplot (figure 6.1), numerically sorted in terms of descending proportions of the *factus est* variant.

While it was observed above that the *factus est* variant is preferred overall, it is notable that the proportional rates of the variants fluctuate strongly between the authors/works and the grand mean for *factus est* is 54.22%. Indeed, just over half of the texts surveyed — Hirtius, Caesar’s *Civ.*, Varro’s *RR*, and the anonymous authors of the *B. Hisp.* and *B. Afr.* — display frequencies of this variant that are a bit less

---

In typological terms, it is worth pointing out that De Sutter’s (2009: 235) Dutch *avc* dataset comprises 789 (33.01%) in the participle-auxiliary order, and 1601 (66.99%) in the auxiliary-participle order. The difference between his dataset and ours is significant: $\chi^2 = 354.98$, $d.f. = 1$, $p < 0.0001$. This shows that the Latin dataset, thus sourced, is practically the reverse of the Dutch counterpart.
than 50%. Furthermore, even between the authors/works which display a share of over 50% of the *factus est* variant, we observe striking fluctuating rates. Overall, there is a highly significant effect of idiolectal differences characterizing the variation in our dataset ($\chi^2 = 120.37$, d.f. = 8, $p < 0.0001$). In more general terms, I note that the difference between texts characterized as “standard” (or “conservative”) versus those which are characterized as “non-standard” (or “innovative”) in their grammatical usage is strikingly borne out by the data. Works that belong in the former category, such as Cicero’s *Phil.*, make heavy use of the diachronically earlier *factus est* variant, whereas half the texts surveyed, which plausibly fit into the latter category (see previous discussions), make use of both variants interchangeably. As we saw above, the grammarians, both ancient and modern, take their paradigms from the Ciceronian word order pattern. It may be that our conception of Latin grammar, as being particularly Cicero-heavy, may need to be modified in light of the above findings. In Chapter 8 the implications of this by-text variability will be discussed in relation to multilevel statistical modelling.
Part III

Information Space
Chapter 7

Methodological Preliminaries

7.1 Introduction

In Part II we examined the structure of the outcome variable of interest — the serialization of the pac — and quantified its overall distribution. We now turn in Part III to identifying what linguistic constraints might, hypothetically, influence the variability in the outcome variable. The set of these hypothetical constraints are here termed the information space (cf. Section 3.3). Throughout Part III, we shall be concerned with a) motivating a host of variables via linguistic theory and empirical evidence, b) operationalizing them in a robust way, and c) quantifying their bivariate effect (i.e. what is their individual effect on the choice of serialization).

The purpose of this chapter is to briefly establish core terminological, methodological, and statistical concepts, which underlie the chapters to come. It is structured as follows. Starting with Section 7.2, I discuss the properties of the information space and how individual variables come to be a part of it. Second, in Section 7.3, I discuss details concerning variable operationalization and some potential pitfalls that one may encounter. Third, Section 7.4 examines the different categories of variables. In Section 7.5, I discuss the statistical concepts that underlie Part III. Section 7.6 adumbrates very briefly how predictor effects might be explained linguistically.

7.2 The information space

7.2.1 Definition

We define a prediction model as \( y \sim \mathbf{x} \), where \( y \) is the outcome variable of interest and \( \mathbf{x} \), where \( \mathbf{x} = x_1, x_2, x_3, \ldots, x_p \), is the set of input variables that \( y \) is taken to be a function of — i.e. the variables that we assume \( y \) to be conditioned upon.\(^1\) Drawing on the econometrics literature (e.g. Davidson & MacKinnon 2004: 17), we further define the set of input variables \( \mathbf{x} \) as being a subset of \( \Omega \), i.e. \( \mathbf{x} \subseteq \Omega \), where \( \Omega \) is termed the information space (or set) of a particular outcome variable of interest. We might thus say that \( \Omega \) is the set of all variables that one can observe on a set of data-points. As Davidson & MacKinnon (2004: 17–18) observe, the variables in the information space are crucially only potential inputs: “Typically, the information set will contain more variables than would actually be used in a [multiple] regression model . . . Such an information set could be very large. As a consequence, much of

\(^1\)For the outcome variable, we find as alternative terms the regressand or response variable. The term “dependent” variable is sometimes used in older textbooks, but its use is discouraged. For the input variable, we alternatively find the covariate, exogenous, explanatory, regressor, or predictor variable used in its place. To confuse matters further, in natural language processing, the term “feature” is often used (Jurafsky & Martin, 2009: 229). The more archaic term “independent” variable is also used, but this is problematic as not all the inputs in a particular model equation may be independent of each other. Consequently, its use is also discouraged.
the art of constructing, or specifying, a regression model is deciding which of the variables that belong to $\Omega$ should be included in the model and which should be excluded’.

7.2.2 Specification of $\Omega$

In order to specify the extent of the information space of the $p$ac alternation, I shall consult a variety of sources and identify variables that have been shown or argued to be important to linguistic variation phenomena. Specifically, I shall draw on theoretical, psycholinguistic, and typological literature to motivate the inclusion of the constraints. I shall also review the evidence of corpus- and socio-variationist studies, to determine how different types of constraints bear upon different alternations across languages. Previous work on the Latin $p$ac alternation is, of course, essential. Further, given that serialization variability in what we have termed AVCS occurs in a number of languages, what do the grammars of these languages and word order studies on their verb cluster have to say about what conditions the variation?

7.3 Operationalizing the predictors

In Part II, we noted that it was essential to code the outcome variable in a way that is empirically robust across trials. Similarly, an important aspect of annotating the dataset for the variables in the information space is their operationalization, that is, “the development of a set of criteria which enables the objective and systematic identification of the values in naturally occurring data” (De Sutter, 2009: 235–236). This is crucial, De Sutter (2009: 227) argues, in order to “[guaranteed] the objectivity, representativity and generalizability of the results”. While research in other social sciences and psychology has long emphasized the accurate measurement of input variables as a matter of high importance, in corpus linguistics research it unfortunately remains wanting (De Sutter, 2009: 237).

Operationalization is more important to some variables than others. For instance, the grammatical notion of number coded as singular and plural measures what it is supposed to measure and is easily replicable, as number is morphosyntactically encoded (e.g. $capt\text{-}us\ est$ ‘he has been captured’ vs. $capt\text{-}i\ sunt$ ‘they have been captured’). The information is indicated by the ending of the participle and the form of the auxiliary verb (at least in finite forms). However, information structure and semantic notions, like aspect, are quite vague and abstract, and, as they stand, are operationally unsafe without detailed coding manuals. One cannot simply posit a variable such as whether the verb is in “focus” or not without (i) defining it accurately and (ii) providing a coding scheme for it in a way that is robust across trials. Only when these two are met can measurement error be minimized. In subsequent chapters (mainly Chapter 9–Chapter 12), we will examine how previous researchers have poorly operationalized predictor variables, as a backdrop for the present operationalizations.

7.4 Types of variables

Variables, whether outcome or inputs, are measured in different ways depending on what exactly it is that you are attempting to tap on a set of observations. For instance, to draw on an example from linguistics, word length and clause type might be two different variables that need to be measured on a set of observations. Intuitively, we need to measure the two in different ways. We might represent the former with real valued numbers from mathematical space $\{1, 2, 3, \ldots\}$ (in e.g. syllables, characters, or phonemes), the latter with a word {main, subordinate}. The system that is used in social science and psychology research stems from Stevens (1946), with some terminological differences. We adopt this system in the present thesis.

---

2Specifically, earlier varieties of German; Dutch; Obolo (Niger-Congo, Cross River); Acholi (Nilo-Saharan, Nilotíć); Ngambay, Mbay (Bongo-Bagirmi, Nilo-Saharan); Breton (Indo-European, Celtic); Tsova-Tush (Nakh-Daghestanian, Nakh); Pipil (Uto-Aztecan, Aztecan); and Purépecha (Tarascan, Tarascan). I thank Matthew Dryer of http://wals.info/ for this information. In practice, though, I shall only discuss German and Dutch.
7.5. STATISTICAL MODELLING

We distinguish between metric (or quantitative) variables and categorical (or qualitative) variables. Both of these have important subtypes.

Metric variables have at base inherent order and “equality of distance” (Stevens, 1946) with respect to the values they take, and as such are measured on real valued number scales. Equality of distance means that, for instance, the difference between 10°C and 30°C is the same as that between 50°C and 70°C, i.e. 20°C in both cases. Examples include the number of clauses per sentence, orthographic words in a clause, and frequency of a word lemma or form.

Categorical variables measure qualitative characteristics on an observation. Some statistics packages, such as R, refer to such variables as factors, and we shall use this terminology interchangeably. Examples include clause type, tense, and sex. We call the values that these variables take levels: to illustrate, the factor sex has two levels, male and female. A distinction is made between ordered factors and unordered factors. Variables such as ratings {very poor, poor, average, good, excellent} have inherent order to them, like metric variables, as for instance poor < good. We might even assign numbers to them: {1, 2, 3, 4, 5}. However, any number assigned to them is fairly arbitrary, and we cannot specify the precise distance between the levels, as we can with metric variables. By contrast, variables such as clause type, lexical category, and finiteness have levels with no inherent order to them. For instance, the variable clause type might involve the levels {main, adjunct, complement, relative}. There is no meaningful way in which a complement clause is “less” than a relative clause. In fact, the set {main, adjunct, complement, relative} is the same as {adjunct, main, relative, complement}. Such variables are thus referred to as unordered factors, though Stevens’ original term nominal is sometimes used. Within unordered factors we distinguish between variables that can only take two levels, such as sex, which we term binary (or dichotomous) variables, and those that take three or more levels, such as clause type, which we term multicategory (or polynomous) variables.

It is essential to grasp the difference between the types of variables. Most importantly, the differences are reflected in different probability distributions for each variable type. For instance, the probability distribution for a binary categorical variable is assumed to follow a binomial distribution, while that for a multicategory variable is assumed to follow a multinomial distribution. As such, statisticians describe different variables with different graphical representations and their differences also affect how they are entered into statistical models, as we shall address shortly. For instance, for ordered categories we need to take account of the ordering to maximize the parsimony of the model, whereas statistical modelling for binary and multicategory responses does not assume dependence on order. The statistical models, as Tutz (2012: 4) calls it, are thus “permutation invariant”.

A final note on typographical conventions. I refer to variable names in teletype font and (for factors) levels in sans serif. For example, the unordered categorical variable SubjectGender, identifying the grammatical gender of the subject, might be associated with three levels, masculine, feminine, and neuter.

7.5 Statistical modelling

Statistical modelling is at the empirical heart of this thesis. In Part III the focus will be on using multilevel generalized linear models, also known as Generalized Linear Mixed Effects Models (GLMERs). Such models are based on Generalized Linear Models (GLMs), themselves based on classical ordinary least squares linear regression (OLS). The purpose of this section is to familiarize the reader with OLS models, before moving onto GLMs and GLMERs. Space does not permit an extensive discussion. For further details, the reader is directed to consult the following textbooks, from which the following material is gleaned: Baayen (2008); Crawley (2012); Draper & Smith (1998); Faraway (2004); Field et al. (2012); Fox (1991, 2002); Gelman & Hill (2007); Harrell (2001); Hutcheson & Sofroniou (1999); Hutcheson & Moutinho (2008); Lewis-Beck (1980); Montgomery et al. (2012); Ott & Longnecker (2010); Tutz (2012); Weisberg (2005).
7.5.1 Ordinary least squares regression

7.5.1.1 Least squares

In simple linear regression, we are interested in predicting the value of the response variable $y$ from a single predictor $x$ by means of a straight line. We write a straight line as the following:

$$y = \beta_0 + \beta_1 x$$  

(7.1)

A simple instantiation of the above equation is given for the dataset in Table 7.1, where we have:

$$y = 0 + 2x$$  

(7.2)

This is expressed graphically in Figure 7.1.

![Figure 7.1: Exact linear relationship: $y = \beta_0 + \beta_1$](image)

Here, $y$ is the response variable, $x$ is the predictor whose information we want to condition on. The terms $\beta_0$ and $\beta_1$ are termed the *regression coefficients* or *parameters*. The parameter $\beta_0$ is termed the *intercept* where the regression line crosses the $y$ axis; in other words, it expresses the value of $y$ when $x = 0$. The parameter $\beta_1$ is termed the *slope* or *gradient*, and this indicates how much $y$ changes given a unit change in $x$. For example, in the figure, when $x = 0$, then $y = 0$, and so $\beta_0 = 0$. When $x$ increases by 1 unit, $y$ increases by 2 units, and thus $\beta_1 = 2$.

The relationship just exemplified is an exact relationship, i.e. we can predict the value of $y$ from a given $x$ precisely (at least from the range of $x$’s we have). However, in linguistics, we do not always come across such precise relationships; often there is a lot of stochastic noise involved in linguistic relationships. We thus need to account for this in our model. We call it the *error term*, denoted $\epsilon$. The equation of the line now looks like:

$$y = \beta_0 + \beta_1 + \epsilon$$  

(7.3)

A linguistically more realistic dataset is given in Table 7.2 and its corresponding relationship is given in Figure 7.2. As can be seen, not all the points are on the regression line, hence the statistical error involved.
<table>
<thead>
<tr>
<th>Observation</th>
<th>$y$</th>
<th>$x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>26</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>28</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>30</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>32</td>
<td>16</td>
</tr>
<tr>
<td>17</td>
<td>34</td>
<td>17</td>
</tr>
<tr>
<td>18</td>
<td>36</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>38</td>
<td>19</td>
</tr>
<tr>
<td>20</td>
<td>40</td>
<td>20</td>
</tr>
<tr>
<td>21</td>
<td>42</td>
<td>21</td>
</tr>
<tr>
<td>22</td>
<td>44</td>
<td>22</td>
</tr>
<tr>
<td>23</td>
<td>46</td>
<td>23</td>
</tr>
<tr>
<td>24</td>
<td>48</td>
<td>24</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>26</td>
<td>52</td>
<td>26</td>
</tr>
<tr>
<td>27</td>
<td>54</td>
<td>27</td>
</tr>
<tr>
<td>28</td>
<td>56</td>
<td>28</td>
</tr>
<tr>
<td>29</td>
<td>58</td>
<td>29</td>
</tr>
<tr>
<td>30</td>
<td>60</td>
<td>30</td>
</tr>
<tr>
<td>31</td>
<td>62</td>
<td>31</td>
</tr>
<tr>
<td>32</td>
<td>64</td>
<td>32</td>
</tr>
<tr>
<td>33</td>
<td>66</td>
<td>33</td>
</tr>
<tr>
<td>34</td>
<td>68</td>
<td>34</td>
</tr>
<tr>
<td>35</td>
<td>70</td>
<td>35</td>
</tr>
<tr>
<td>36</td>
<td>72</td>
<td>36</td>
</tr>
<tr>
<td>37</td>
<td>74</td>
<td>37</td>
</tr>
<tr>
<td>38</td>
<td>76</td>
<td>38</td>
</tr>
<tr>
<td>39</td>
<td>78</td>
<td>39</td>
</tr>
<tr>
<td>40</td>
<td>80</td>
<td>40</td>
</tr>
<tr>
<td>41</td>
<td>82</td>
<td>41</td>
</tr>
<tr>
<td>42</td>
<td>84</td>
<td>42</td>
</tr>
<tr>
<td>43</td>
<td>86</td>
<td>43</td>
</tr>
<tr>
<td>44</td>
<td>88</td>
<td>44</td>
</tr>
<tr>
<td>45</td>
<td>90</td>
<td>45</td>
</tr>
<tr>
<td>46</td>
<td>92</td>
<td>46</td>
</tr>
<tr>
<td>47</td>
<td>94</td>
<td>47</td>
</tr>
<tr>
<td>48</td>
<td>96</td>
<td>48</td>
</tr>
<tr>
<td>49</td>
<td>98</td>
<td>49</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 7.1: Dataset providing an exact linear relationship
<table>
<thead>
<tr>
<th>Observation</th>
<th>$y$</th>
<th>$x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.76</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>4.41</td>
<td>0.58</td>
</tr>
<tr>
<td>3</td>
<td>7.43</td>
<td>0.71</td>
</tr>
<tr>
<td>4</td>
<td>7.58</td>
<td>0.93</td>
</tr>
<tr>
<td>5</td>
<td>8.24</td>
<td>0.96</td>
</tr>
<tr>
<td>6</td>
<td>8.98</td>
<td>0.98</td>
</tr>
<tr>
<td>7</td>
<td>10.78</td>
<td>1.23</td>
</tr>
<tr>
<td>8</td>
<td>10.79</td>
<td>1.36</td>
</tr>
<tr>
<td>9</td>
<td>14.53</td>
<td>1.56</td>
</tr>
<tr>
<td>10</td>
<td>17.62</td>
<td>1.66</td>
</tr>
<tr>
<td>11</td>
<td>18.55</td>
<td>1.75</td>
</tr>
<tr>
<td>12</td>
<td>22.88</td>
<td>2.11</td>
</tr>
<tr>
<td>13</td>
<td>23.05</td>
<td>2.18</td>
</tr>
<tr>
<td>14</td>
<td>25.65</td>
<td>2.49</td>
</tr>
<tr>
<td>15</td>
<td>28.43</td>
<td>2.52</td>
</tr>
<tr>
<td>16</td>
<td>29.30</td>
<td>2.75</td>
</tr>
<tr>
<td>17</td>
<td>30.03</td>
<td>2.84</td>
</tr>
<tr>
<td>18</td>
<td>32.28</td>
<td>2.86</td>
</tr>
<tr>
<td>19</td>
<td>33.74</td>
<td>2.87</td>
</tr>
<tr>
<td>20</td>
<td>36.95</td>
<td>3.14</td>
</tr>
<tr>
<td>21</td>
<td>37.36</td>
<td>3.22</td>
</tr>
<tr>
<td>22</td>
<td>37.39</td>
<td>3.65</td>
</tr>
<tr>
<td>23</td>
<td>37.90</td>
<td>3.93</td>
</tr>
<tr>
<td>24</td>
<td>39.19</td>
<td>4.08</td>
</tr>
<tr>
<td>25</td>
<td>45.93</td>
<td>4.18</td>
</tr>
<tr>
<td>26</td>
<td>47.37</td>
<td>4.63</td>
</tr>
<tr>
<td>27</td>
<td>47.67</td>
<td>4.70</td>
</tr>
<tr>
<td>28</td>
<td>52.63</td>
<td>4.93</td>
</tr>
<tr>
<td>29</td>
<td>54.74</td>
<td>5.19</td>
</tr>
<tr>
<td>30</td>
<td>58.14</td>
<td>5.39</td>
</tr>
<tr>
<td>31</td>
<td>61.57</td>
<td>5.58</td>
</tr>
<tr>
<td>32</td>
<td>61.90</td>
<td>6.14</td>
</tr>
<tr>
<td>33</td>
<td>62.84</td>
<td>6.34</td>
</tr>
<tr>
<td>34</td>
<td>66.82</td>
<td>6.86</td>
</tr>
<tr>
<td>35</td>
<td>69.38</td>
<td>6.99</td>
</tr>
<tr>
<td>36</td>
<td>70.46</td>
<td>7.08</td>
</tr>
<tr>
<td>37</td>
<td>71.65</td>
<td>7.60</td>
</tr>
<tr>
<td>38</td>
<td>73.69</td>
<td>7.62</td>
</tr>
<tr>
<td>39</td>
<td>76.84</td>
<td>8.03</td>
</tr>
<tr>
<td>40</td>
<td>76.89</td>
<td>8.45</td>
</tr>
<tr>
<td>41</td>
<td>77.09</td>
<td>8.57</td>
</tr>
<tr>
<td>42</td>
<td>77.45</td>
<td>8.57</td>
</tr>
<tr>
<td>43</td>
<td>78.33</td>
<td>8.74</td>
</tr>
<tr>
<td>44</td>
<td>79.43</td>
<td>8.75</td>
</tr>
<tr>
<td>45</td>
<td>86.55</td>
<td>8.78</td>
</tr>
<tr>
<td>46</td>
<td>87.44</td>
<td>9.13</td>
</tr>
<tr>
<td>47</td>
<td>87.85</td>
<td>9.64</td>
</tr>
<tr>
<td>48</td>
<td>91.28</td>
<td>9.67</td>
</tr>
<tr>
<td>49</td>
<td>95.51</td>
<td>9.80</td>
</tr>
<tr>
<td>50</td>
<td>96.67</td>
<td>9.99</td>
</tr>
</tbody>
</table>

Table 7.2: Dataset providing an inexact linear relationship
In order to find the regression line that “best fits” the data, we use the method of least squares, that is by finding the line that minimizes the sum of the squared errors. For any line that is drawn on a dataset which is not perfectly predictable, the values will be at some distance from that particular line. You will notice on the Figure 7.2 that while some points fall on the line precisely, others are further away from it. We define the error of a \( y \) value as the distance of the observed \( y \) from the \( \hat{y} \) predicted by the line (the “hat” character indicates a predicted value). We specify it with the following equation:

\[
\epsilon_i = y_i - \hat{y}_i
\]  

Thus, we might want to add all of these errors up to determine the total sum of errors for a particular regression line, \( \sum_{i=1}^{n} \epsilon_i \). However, this is problematic, as the points are scattered below the regression line (points the regression line overpredicts) which result in negative errors and points scattered above the regression line (points that the regression line underpredicts) which result in positive errors. Unfortunately, adding negatives and positives together cancel each other out, so we can’t use this method. Consequently, the method that we use is to sum the squared errors (\( SSE \)). It is specified by the following equation:

\[
SSE = \sum_{i=1}^{n} \epsilon_i^2
\]  

Unfortunately, it is rather tedious to compute by hand the sums of the squared errors for every line imaginable. There are potentially an infinite number of lines that we might wish to assess and compare. So, we turn to algebra. First of all, we need five key values, Crawley’s (2012) “Famous Five”:

1. \( \sum_{i=1}^{n} y_i \) — the sum of the \( y \)'s
2. \( \sum_{i=1}^{n} y_i^2 \) — the sum of the squared \( y \)'s
3. \( \sum_{i=1}^{n} x_i \) — the sum of the \( x \)'s

\[(7.5)\]
CHAPTER 7. METHODOLOGICAL PRELIMINARIES

\[
\begin{align*}
\sum_{i=1}^{n} y_i & \quad \sum_{i=1}^{n} y_i^2 & \quad \sum_{i=1}^{n} x_i & \quad \sum_{i=1}^{n} x_i^2 & \quad \sum_{i=1}^{n} x_i y_i \\
2379.918 & \quad 152975.7 & \quad 241.788 & \quad 1625.659 & \quad 15739.1
\end{align*}
\]

Table 7.3: Crawley’s (2012) Famous Five for our fictional dataset

4. \( \sum_{i=1}^{n} x_i^2 \) — the sum of the \( x \)'s squared
5. \( \sum_{i=1}^{n} x_i y_i \) — the sum of the products

The Famous Five for our fictional dataset are reported in Table 7.3. Using these five formulae, we then compute the three corrected sums for \( x \), \( y \) and the product term \( xy \), which takes into account the number of observations in the dataset \( n \):

1. \( SSX = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} \)

2. \( SSY = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \)

3. \( SSXY = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n} \)

Note that \( n \) = number of observations. We now compute \( SSX, SSY \) and \( SSXY \) for our fictional dataset.

1. \( SSX = 1625.659 - \frac{(241.788)^2}{50} = 456.4303 \)
2. \( SSY = 152975.7 - \frac{(2379.918)^2}{50} = 39695.51 \)
3. \( SSXY = 782 - \frac{(55)(111)}{50} = 4230.388 \)

We then use this information to calculate the slope \( \beta_1 \) and the intercept \( \beta_0 \). The slope is computed as follows:

\[
\beta_1 = \frac{SSXY}{SSX} \quad (7.6)
\]

Thus, for our dataset, this is equivalent to:

\[
\beta_1 = \frac{SSXY}{SSX} = \frac{4230.388}{456.4303} = 9.268421 \quad (7.7)
\]

The intercept is calculated thus:

\[
\beta_0 = \bar{y} - \beta_1 x = \frac{\sum y}{n} - \beta_1 \frac{\sum x}{n} = \bar{y} - \beta_1 \bar{x} \quad (7.8)
\]

And for our dataset this is equivalent to:

\[
\beta_0 = 2.778491 = 47.59836 - 9.268421 \times 4.835761 \quad (7.9)
\]
7.5. STATISTICAL MODELLING

Our model equation computed via ordinary least squares is as follows:

\[
\hat{y} = \beta_0 + \beta_1 x = 2.778491 + 9.268421 x \tag{7.10}
\]

This says that when \( x = 0 \), the average value of \( y \) is 2.78 (the intercept or \( \beta_0 \)) and that for a one unit increase in \( x \) then \( y \) increases on average by 9.27 units (the slope or \( \beta_1 \)). The regression line can then be plotted (as it has been done in Figure 7.2). However, there are a quite a few things that need to be done first.

7.5.1.2 Statistical inference

Given that we have used a sample to estimate our model parameters, we need to know whether they are robust enough to apply confidently to the population from which the sample was drawn. The true population parameters, which we denote as \( \hat{\beta}_1 \) and \( \hat{\beta}_0 \) are unknown quantities and had to be estimated from the sample. However, can it indeed be claimed that the intercept and the slope we have sampled is reflective of what it might be in the population from which the sample was drawn? Might they reflect the “true” population values poorly? Might the population values in fact be zero in the population? We need to assess this. Several terms are important here.

7.5.1.2.1 Hypothesis testing

When we evaluate the parameters, we construct two hypotheses: the null hypothesis \( H_0 \) and the alternative hypothesis \( H_1 \). For instance, the null hypothesis asserts that, in the case of \( \hat{\beta}_1 \), it is zero in the population, while the alternative hypothesis is that, in the case of \( \hat{\beta}_1 \), it is not zero in the population.

7.5.1.2.2 Error variance

The error variance, denoted \( s^2 \), is the error sum of squares divided by the number of degrees of freedom (i.e., the number of observations minus the number of parameters being estimated; in this case, \( n - 2 \)). For our dataset this is:

\[
s^2 = \frac{SSE}{(n - 2)} = \frac{486.4758}{50 - 2} = \frac{486.4758}{48} = 10.13491 \tag{7.11}
\]

7.5.1.2.3 Standard error

Essential values are the standard errors of the slope and the intercept. The standard error is an estimate of the standard deviation of these estimates. The standard error of the slope is denoted \( se_{\beta_1} \) and that of the intercept \( se_{\beta_0} \), and are calculated as follows:

\[
se_{\beta_1} = \sqrt{s^2 SSX} \tag{7.12}
\]

\[
se_{\beta_0} = \sqrt{s^2 \sum x^2 \over n \times SSX} \tag{7.13}
\]

For our dataset, these are computed thus:

\[
se_{\beta_1} = \sqrt{10.13491 \over 456.4303} = 0.1490125 \tag{7.14}
\]

\[
se_{\beta_0} = \sqrt{10.13491 \times 1625.659 \over 50 \times 456.4303} = 0.8496744 \tag{7.15}
\]

7.5.1.2.4 Confidence intervals

Using the standard errors, we can construct confidence intervals for the population regression coefficients, that is we can specify a range where we might expect these parameters to fall in between. In other words, it specifies the inaccuracy in the parameter estimates. Ideally, we
want tight confidence intervals with a small range. We typically use the 95% confidence interval, which means that in 95% of samples, the estimates should be expected to fall within that range. Confidence intervals that include zero indicate that the true population regression coefficient could well be zero. If the confidence interval does not include zero, then it can be concluded with 95% confidence that the sample parameter estimate is reflective of the population parameter estimate. Generally:

$$\beta \pm t_{n-2;0.975}se\beta$$  \hspace{1cm} (7.16)

What does this mean? First, we draw on the t-distribution and examine the critical t-value for 50 − 2 degrees of freedom at the 95% confidence interval. We see that this is 2.010635 (for greater values of degrees of freedom, it is 1.96). We can then calculate the confidence intervals:

$$\beta_1 \pm 2.010635 \times 0.1490125 = 0.2996097$$  \hspace{1cm} (7.17)

$$\beta_0 \pm 2.010635 \times 0.8496744 = 1.708385$$  \hspace{1cm} (7.18)

For $\beta_1$, they range from 8.968811 (9.268421 − 0.2996097) to 9.568031 (9.268421 + 0.2996097). These confidence intervals do not include zero, and we can therefore conclude with 95% certainty that the sampled parameter $\beta_1$ is reflective of the population slope. Additionally, for the intercept $\beta_0$, they range from 1.070106 (2.778491 − 1.708385) to 4.486876 (2.778491 + 1.708385), thus indicating that the population intercept is reflected by the sample intercept coefficient.

7.5.1.3 Regression diagnostics

Classical linear regression makes specific assumptions about the structure of the data. Specifically, it assumes, amongst other things:

1. Normality of errors;
2. Linearity.

Consequently, it needs to be explored whether the data violate these assumptions. If so, the estimates may be biased and therefore unreliable, and other regression methods should be used. This is a key point which will be returned to when I discuss the modelling of binary responses, such as that of the pac alternation, below.

7.5.1.3.1 Normality of errors  Linear regression is most robust when the residuals (the errors) are normally distributed with a mean of zero and a standard deviation of 1.

7.5.1.3.2 Linearity  As the term implies, linear regression concerns fitting a straight line. If the relationship between the predictor $x$ and $y$ is essentially non-linear, then linear regression cannot be used. As can be gleaned from Figure 7.2, the relationship we have here is clearly of a linear nature.

7.5.1.4 Dummy variable coding for categorical predictors

The focus thus far has been on numerical variables, predicting a numerical response $y$ as a function of a numerical predictor $x$, but a crucial aspect of this investigation concerns categorical data, as discussed above. We therefore need to see how it is modelled in a classical linear regression, before we can understand how it is modelled in more complex types of regression.

3Note that several assumptions are omitted here.
7.5.1.4.1 Binary variables  Binary categorical variables cannot be entered directly into a linear regression analysis, but it is necessary to code them as “dummy variables”. The method for binary categorical variables is that of treatment coding (also: indicator coding), whereby one level of the factor is coded as 0 and another factor is coded as 1. For instance, one could recode the binary variable `Number`, with the two levels `singular` and `plural`, with `singular` coded as 0 and `plural` coded as 1. Just like for a metric variable, where the value of $\beta_0$ indicates the average value of $y$ when $x = 0$ and $\beta_1$ indicates the average increase in $y$ for a one unit change in $x$, so a binary categorical is interpreted: the value of $\beta_0$ indicates the average value of $y$ when the binary variable is 0 and $\beta_1$ indicates the average increase in $y$ when the level coded 0 is changed to the level coded as 1. In other terms, the level coded 0 acts as the “baseline” or “reference” category to which the level coded 1 is compared. It is important to note that not all levels are entered into a regression model. For instance, we do not enter simultaneously into a model a variable for `Number1` with `singular` coded as 0 and `plural` coded as 1 and a variable for `Number2` with `singular` coded as 1 and `plural` coded as 0. Specifically, if there are $j$ levels in a factor, $j - 1$ dummy variables are used in the regression analysis to avoid perfect collinearity.

To exemplify how this works, a binary variable is added to the fictional dataset, termed `j` with the levels `j1` and `j2`, with `j1` coded 0 and `j2` coded 1. This is plotted in Figure 7.3, along with a regression line. We will do the regression in R now:

Call:
\[ \text{lm(formula = y \sim j, data = reg)} \]

Residuals:
\[
\text{Min} \quad -25.0376 \quad -11.7094 \quad 0.7712 \quad 10.5850 \quad 24.2624 \\
\text{1Q} \quad -10.7077 \quad -2.3756 \quad 0.0 \quad 2.3756 \quad 8.3704 \\
\text{Median} \quad 0.7712 \quad 0.7712 \quad 0.7712 \quad 0.7712 \quad 0.7712 \\
\text{3Q} \quad 10.5850 \quad 10.5850 \quad 10.5850 \quad 10.5850 \quad 10.5850 \\
\text{Max} \quad 24.2624 \quad 24.2624 \quad 24.2624 \quad 24.2624 \quad 24.2624 \\
\]

Coefficients:
\[
\text{Estimate} \quad \text{Std. Error} \quad t \text{ value} \quad \text{Pr(>|t|)} \\
(\text{Intercept}) \quad 22.790 \quad 2.727 \quad 8.357 \quad 6.39e-11 \quad *** \\
\text{j} \quad 49.618 \quad 3.856 \quad 12.866 \quad < 2e-16 \quad *** \\
---

Figure 7.3: Binary variable
The intercept says that when the variable \( j \) is \( j_1 \), that is when it is 0, then the average value of \( y \) is 22.790. The standard error associated with this estimate is 2.727, which going on the sample size and 48 degrees of freedom means that the confidence intervals for the estimate are \([17.3071, 28.2729]\). The \( t \)-statistic is \( 22.790/2.727 \), which on 48 degrees of freedom is much higher than the critical \( t \)-value of 2.01, and so we can conclude that the parameter is significantly different from 0. When the variable is changed from the level \( j_1 \) to \( j_2 \), that is for a unit increase in the value of \( x \) from 0 to 1, the value of \( y \) increases on average by 49.618 units, which can be clearly discerned from the regression plot. The standard error associated with this estimate is 3.856, which results in confidence intervals of \([41.8636, 57.3716]\). The \( t \)-value, \( 12.866 = 49.618/3.856 \), indicates that we can have 95% confidence that the estimate is reflective of the population estimate. In the output above, I deliberately coded \( j_1 \) as 0 and \( j_2 \) as 1 manually, that is as a metric variable. However, \( R \) is clever and automatically converts factor levels into (0-1)-coding. For instance, if I had coded these as a factor and let \( R \) do the work, the same regression coefficients would have been obtained:

```
Call:
  lm(formula = y ~ as.factor(j), data = reg)

Residuals:
     Min      1Q  Median      3Q     Max
-25.0376 -11.7094  0.7712  10.5850  24.2624

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  22.790     2.727   8.357  6.39e-11 ***
as.factor(j)1  49.618     3.856  12.866  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: 13.63 on 48 degrees of freedom
Multiple R-squared:  0.7752,  Adjusted R-squared:  0.7705
F-statistic: 165.5 on 1 and 48 DF,  p-value: < 2.2e-16
```

7.5.1.4.2 Multicategory variables

Often the factor has more than two levels. For multicategory factors, there are several options available, depending on whether they are unordered or ordered. Ordered factors are not used in this thesis, so we will concentrate on unordered ones. Here, we shall take the three-level categorical variable \( k \) with three levels \( k_1 \), \( k_2 \), and \( k_3 \).

#### 7.5.1.4.2.1 Treatment coding

For unordered multicategory variables, one way is to follow that used for binary variables and use a treatment coding method, using \( j - 1 \) dummy variables. Thus, to avoid perfect collinearity, two predictors are entered into the model. Here, \( k_1 \) is coded as the baseline to which \( k_2 \) and \( k_3 \) are to be compared. A first predictor, call it \( k_a \), takes two values, one indicating the presence of \( k_2 \) (= 1) and one indicating the absence of \( k_2 \) (= 0), that is, when the level is \( k_1 \) or \( k_3 \). A second predictor, call it \( k_b \), takes two values, one indicating the presence of \( k_3 \) (= 1) and one indicating the absence of \( k_3 \) (= 0), that is, when the level is \( k_1 \) or \( k_2 \). Again, there is no term for the presence or
absence of \( k_1 \), as it is perfectly predictable via the conjunction of the terms \( k_a \) and \( k_b \), and would thus yield perfect collinearity. The regression model is as follows:

**Call:**
```r
lm(formula = y ~ k, data = reg)
```

**Residuals:**
```
             Min      1Q     Median      3Q        Max
-16.419  -6.381   -1.818   7.640    19.769
```

**Coefficients:**
```
               Estimate  Std. Error t value   Pr(>|t|)
(Intercept)    12.946      2.581   5.016 7.97e-06 ***
ka             28.855      3.485   8.279 9.80e-11 ***
k_b            65.373      3.402  19.219 < 2e-16 ***
---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
```

Residual standard error: 9.657 on 47 degrees of freedom
Multiple R-squared: 0.8896, Adjusted R-squared: 0.8849
F-statistic: 189.3 on 2 and 47 DF, p-value: < 2.2e-16

Note that there are 47 degrees of freedom because we have one extra parameter in the model. Note that it is possible to see whether \( k_2 \) and \( k_3 \) are significantly different as to their effect on \( y \) by changing the reference category. However, as will be explained below, this is not advisable as it runs the risk of multiple comparisons and Type I error, that is when the null hypothesis is anti-conservatively rejected.

### 7.5.1.4.2.2 Sum coding

An alternative scheme — which is the default dummy coding method in some statistical software — is that of sum (or deviation) coding. In this type of dummy coding, a variable is coded numerically for three values, that is \(-1, 0, 1\). Again, \( j - 1 \) dummy variables are entered into the model to prevent perfect collinearity. The level that is left out of the model is coded as \(-1\), the last level in the factor, the level of interest is coded as \(1\), and the other levels are coded as \(0\). Here, \( k_1 \) is coded as the level which is to be left out of the analysis. A first predictor, call it \( k_a \), takes three values, one indicating the level left out (= \(-1\)), one indicating the presence of \( k_2 \) (= \(1\)) and one indicating the absence of \( k_2 \) (= \(0\)), that is, when the level is \( k_3 \). A second predictor, call it \( k_b \), takes three values, one indicating the level left out (= \(-1\)), one indicating the presence of \( k_3 \) (= \(1\)) and one indicating the absence of \( k_3 \) (= \(0\)), that is, when the level is \( k_2 \). Again, no term is included for the presence or absence of \( k_1 \), as it is perfectly predictable via the conjunction of the terms \( k_a \) and \( k_b \), and would thus yield perfect collinearity. The regression model is as follows:

**Call:**
```r
lm(formula = y ~ ka, data = reg)
```

**Residuals:**
```
            Min      1Q     Median      3Q        Max
-16.419   -6.381    -1.818    7.640    19.769
```

**Coefficients:**
```
               Estimate   Std. Error t value   Pr(>|t|)
(Intercept)    44.356       1.377   32.220  <2e-16 ***
```

71
The output here means the following. First, the intercept, rather than meaning the mean of the reference group (which doesn’t exist), refers to the grand mean value of $y$ for all the categories of the $x$ variable: $44.355 = (12.94643+41.80118+78.31947)/3$. The slopes now indicate the deviation of each of the specified factor levels from this grand mean. Thus, the average for the level $k1$ is $12.94643$, which is $31.409$ units less than the grand mean $44.356$, and this is a significant deviation ($-31.409/2.029 = -15.482$) The average for the level $k2$ is $41.80118$, which is $-2.555$ less than the grand mean, a coefficient estimate that is not significant. Note that only the regression coefficients are affected by this change in coding method; the model fit remains the same, for instance.

In this thesis, however, treatment coding will be used, as it is the most interpretable (so long as one selects the largest category as the baseline) and is more applicable to unbalanced datasets (see e.g. Baayen 2008).

### 7.5.2 Generalized linear models

#### 7.5.2.1 Introduction to binary response models

While responses are often metric, they can also be categorical whereby they take two or more levels. For instance, in epidemiology, one might wish to consider whether a patient dies or survives as a result of exposure to a particular treatment; in social housing, one might wish to consider whether house ownership is significantly influenced by marriage, gender, and age, as predictors. Turning to linguistics, we are usually interested in predicting whether a certain type of grammatical structure is more likely given a certain set of predictors rather than another grammatical structure. This is precisely the case with the PAC alternation, where the response is a binary choice between *factus est* and *est factus*. Consequently, it is essential to know how these types of binary responses are modelled.

The data we will consider here as an illustration of modelling a binary response variable has been taken from the freely available dataset *dative* from the *languageR* package in R (Baayen, 2013), a dataset which concerns the English dative alternation and is discussed in detail in Bresnan et al. (2007). The response variable is whether a speaker chooses the prepositional dative as in (1a) or the double-object construction as in (1b).

(1) **Dative alternation**

- a. John gave the book **to Mary** ← PREPOSITIONAL-DATIVE  
  THEME    RECIPIENT
- b. John gave **Mary** the book ← DOUBLE-OBJECT  
  RECIPIENT    THEME

I will refer to the PREPOSITIONAL-DATIVE structure as the PP-outcome and the DOUBLE-OBJECT structure as the NP-outcome. As a putative predictor we consider here the length of the recipient (in words), a metric predictor.
7.5.2.2 Modelling binary data in linear regression

One possibility that could be explored is fitting a linear regression to the data, whereby the response variable is dummy coded via (0-1)-coding. In this case one of the levels of the factor represents a success (the one coded 1) and the other level of the factor represents a failure (the one coded 0). For the dative data, we will treat the PP-outcome as a success and a NP-outcome as a failure. There are 3263 observations in the dative dataset, with 849 (26%) PP-outcomes and 2414 (74%) NP-outcomes. We thus have a vector of 848 1’s and 2414 0’s.

If the raw data is plotted as a function of the metric predictor concerning the length of the recipient (cf. Figure 7.4), we obtain a plot whereby all the NP-outcomes are on the 0 value of the y axis and all the PP-outcomes are on the 1 value of the y axis (this is indicated by the unfilled black circles). Each point represents a single observation. Given that there are more successes for longer recipients, it appears that the longer the recipient, the more likely a PP-outcome. However, because the values can only take on 0 or 1, and are thus restricted, the overall relationship is unclear. Another way to show this therefore is to illustrate for each value (or binned values) of the length of the recipient the proportion of observations that are realized as a prepositional recipient. These are illustrated in Figure 7.4 with blue filled circles. This indicates, for instance, that if the length of the recipient is of 1 word in length, then the proportion of a prepositional phrase recipient is 0.12, whereas if the length of recipient is of 5 words in length, then the proportion of a prepositional phrase recipient is 0.76. A regression line is also plotted, following the output of a linear regression analysis below. The regression output then says that when \( x = 0 \), the probability of a prepositional variant is 0.1,\(^4\) and for a one unit increase in the length of the recipient, the probability estimate increases by 0.09. When the estimates \( \beta_0 \) and \( \beta_1 \) are divided by their standard errors, significant \( t \)-statistics are obtained, which are above the critical value of 1.96 for a model with 3261 degrees of freedom.

Call:

\( \text{lm(formula = Response} \sim \text{LengthOfRecipient, data = dative)} \)

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
</table>

\(^4\)Meaningless in this instance.
-1.85451 -0.18526 -0.18526 0.01397 0.81474

Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|---------|
| (Intercept) | 0.096283 | 0.009342 | 10.31 <2e-16 *** |
| LengthOfRecipient | 0.088975 | 0.003374 | 26.37 <2e-16 *** |

---

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.3984 on 3261 degrees of freedom
Multiple R-squared: 0.1758, Adjusted R-squared: 0.1755
F-statistic: 695.5 on 1 and 3261 DF, p-value: < 2.2e-16

While the above seems a fairly straightforward application of linear modelling to binary response data, there are several problems with such an analysis. Linear regression assumes normality of the errors $\epsilon_i$ with $\mathcal{N}(0, \sigma^2)$. The diagnostic plots for normality as displayed in Figure 7.5 show that this is clearly not the case. For binary data, the error can never be normally distributed.

![Diagnostic plots](image)

Figure 7.5: Diagnostic plots indicating problems of modelling binary data in OLS regression

The assumption of linearity for linear regression is also not met. For instance, rather than clustering around the regression line, the proportions for each value of the $x$ variable (the filled blue circles) show more of a curve. This kurtosis cannot be adequately modelled via linear regression, as it results in a poor reflection of the data. Impressionistically, one can also see that it’s not a particularly great

---

5We will see below that the predictor is in fact unsuitable for including untransformed into the models discussed below.
fit to the data. Third, in terms of variance, there is a clear heteroskedastic pattern. According to
the ncvTest in the car package, this pattern is highly significant, indicating non-constant variance
($\chi^2 = 191.3589, DF = 1, p = 1.605921 \times 10^{-43}$). Most crucially, probabilities lie in the range $[0, 1]$. It will
be noted that the regression line actually predicts probabilities that are outside the realms of probability
space. For instance, the “probability” of a prepositional phrase outcome when the recipient is 25 words
in length is $2.320658 = 0.096283 + (0.088975 \times 25)$. Similarly, the “probability” of a prepositional phrase
outcome when the recipient is 30 words in length is $2.765533 = 0.096283 + (0.088975 \times 30)$. These values
are mathematically meaningless. With binary categorical outcomes, the mean value of $y$ must lie between
0 and 1. Values outside this range are impossible.

7.5.2.3 Logit link function

In order to get around these problems, it is necessary to use a generalized version of the linear model,
whereby an attempt is made to render the association between the predictor and the response linear.
Generalized forms of the linear model (the ‘Generalized Linear Model’ or glm) have the form:

$$g(\mu(y)) = \eta = \beta_0 + \beta_1 x_1 + \epsilon$$ (7.19)

where $\beta_0 + \beta_1 x_1$, i.e. $\eta$, is the linear predictor (the systematic component), $y$ is the response variable
(the random component), $g()$ is the link function that relates the mean of the random component to the
systematic component via a specified transformation, and $\epsilon$ is the model’s error structure.

The link function is chosen such that it optimizes the relationship between the random component
and the systematic component and such that the predicted values of $y$ are mathematically possible. For
classical linear regression, which can also be embedded into a glm framework, the link function is one
of identity: the mean of $y$ is equal to the systematic component. Link functions come with their own
assumptions; for instance, with the identity link, it assumed that the data have normally distributed
errors and the relationship is one of a direct linear nature. For binary response models, the link function
cannot be one of identity as that implies classical linear regression, which, as we have already seen, is not
an accurate modelling of the data. A transformation on the mean value of $y$ given a specific value of $x$
needs to take place so that impossible probabilities do not result. The link made use of here is the logit
(or logistic or log of the odds) link function:

$$\log \left( \frac{\pi(\eta)}{1 - \pi(\eta)} \right)$$ (7.20)

This means that the odds are taken, and then the natural logarithm of the odds is applied. There are
a number of reasons why this is useful. Probabilities are constrained between 0 and 1. Odds are like
probabilities in that they have a lower limit of 0, but they have no upper limit, that is, they extend
towards infinity. By transforming probabilities into odds and then back into probabilities again, it means
that the maximum bound of 1 is never exceeded. Taking the log of the odds solves the problem of
probabilities being bounded by zero on their lower limit. The natural logarithm of a number extends
from negative infinity to zero. When converted back into probabilities, they lie in mathematically possible
space. When the logit link function is applied, the resulting statistical model is often known as a logistic
regression.

7.5.2.4 Maximum Likelihood Estimation

In classical linear regression, the parameters $\beta_0$ and $\beta_1$ are estimated by ordinary least squares, as we
have seen, that is by finding the line that minimizes the sum of the squared deviation. For glmS, of any
type, the procedure used to estimate the parameters is known as maximum likelihood estimation, that is

\(^6\)Even if this involves transforming one or more of the $x$ variables.
the parameters are found such that they make the observed patterns most probable based on the sample data. We will not discuss the technicalities of this here, but the interested reader is directed to Eliason (1993) for discussion.

7.5.2.5 An example

We will now fit a glm in order to see how to interpret an example model. The model output is shown below for a glm of the dative response as a function of length of the recipient.

\[
\logit[P(y = 1|x)] = \beta_0 + \beta_x
\]  

(7.21)

The logit probability of \( y \) given a specific value of the length of the recipient is a function of the linear predictor.

Call:

```
glm(formula = RealizationOfRecipient ~ LengthOfRecipient, family = binomial, 
data = dative)
```

Deviance Residuals:

<table>
<thead>
<tr>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.3757</td>
<td>-0.5954</td>
<td>-0.5954</td>
<td>0.1368</td>
<td>1.9066</td>
</tr>
</tbody>
</table>

Coefficients:

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | -2.34116 | 0.07951 | -29.44 | <2e-16 *** |
| LengthOfRecipient | 0.70085 | 0.03866 | 18.13 | <2e-16 *** |

---

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3741.1 on 3262 degrees of freedom
Residual deviance: 3104.9 on 3261 degrees of freedom
AIC: 3108.9

Number of Fisher Scoring iterations: 5

There are two main aspects to interpreting the output of glm. The first is an interpretation of the model fit, the second is an interpretation of the parameters.

In order to determine whether a model is useful and an accurate reflection of the data a number of model fit statistics are assessed. Here, the properties of a null model (that is a model with only the intercept in it) is compared with the fuller model (the model with the intercept and the predictor in it).

Log-likelihood and deviance (\(-2LL\)). In order to assess the fit of a model, log-likelihood and deviance statistics are used. The log-likelihood is a measure of how well the fitted values reflect the observed values, and is comparable in some respects to the residual sum of squares for classical linear regression. The log-likelihood is usually presented as a deviance statistic, which is \(-2\) times the log-likelihood, and has a \(\chi^2\) distribution. For the model above, the “null deviance” indicates how well the model with just the intercept \(\beta_0\) term fits the data: as we can see, this is 3741.1 on 3262 degrees of freedom (the degrees of freedom are the number of observations \(n\) minus the number of parameters being evaluated, i.e. 1 for the intercept). A \(-2LL\) value of 0 means that there is no deviance between the observed and fitted values. The “residual deviance” indicates how well the model with both the intercept \(\beta_0\) and the predictor
7.5. STATISTICAL MODELLING

coefficient $\beta_1$ fits the data. A $-2LL$ value of 0 again indicates that there is no deviance between the observed and fitted values. Given that $-2LL$ follows a $\chi^2$ distribution, it should follow that in an ideal world the null deviance should be significant, and the residual deviance non-significant. In other words, we do not want to see significant deviance, as that means there is a lot of variance left to be accounted for. However, this does not usually happen. What is more interesting is comparing whether the difference between the null deviance, for the model with just the intercept, and the residual deviance, for the model including the intercept and the predictor coefficient of interest, is significantly large enough.

$$-2LL_{\text{difference}} = (-2LL_{\text{null}}) - (-2LL_{\text{residual}})$$ (7.22)

For this model, the difference is significantly large enough $3741.1 - 3104.9 = 636.2$, and we can thus conclude that adding the predictor significantly reduces the deviance. (The number of degrees of freedom on which the $-2LL$ should be evaluated is the difference in $df$ of the null model and the parameterized model.)

In classical linear regression, we used a $R^2$ (or adjusted $R^2$) value to determine how much variance in the response variable was accounted for by the inclusion of (a) specific predictor(s). The $R^2$ statistic cannot be directly applied to logistic regression models, and instead several pseudo-$R^2$ statistics are usually reported. The one discussed here is that of Nagelkerke (1991), and is computed as follows:

$$\text{pseudo-}R^2 = \frac{1 - \exp\left(\frac{(-2LL(m2)) - (-2LL(m1))}{n}\right)}{1 - \exp\left(\frac{-2LL(m1)}{n}\right)}$$ (7.23)

This measure will feature infrequently in this thesis, because it is irrelevant in multilevel models.

A further way to assess the fit of the model is to examine various information criteria in terms of its model parsimony, such as the Akaike Information Criterion ($AIC$) (Akaike, 1974). The $AIC$ of the null model is compared to that of the fuller model with the predictor term in it. The model is chosen which has values closest to zero. The $AIC$ is calculated as:

$$AIC = -2LL + 2k$$ (7.24)

After the model has been evaluated, we may wish to interpret the parameters in the model. The interpretation of the parameters is fairly straightforward. The estimate and the standard error (for GLMS, the asymptotic standard error) are interpreted in much the same way as in classical linear regression, as already discussed.

For the intercept, the difference is that, while in classical linear regression, when $x = 0$, the average value of $y$ is $\beta_0$, in logistic regression models, when $x = 0$, the logit probability of $y$ is $\beta_0$. For the slope, the difference is that, while in classical linear regression, a one-unit change in the value of the predictor variable results in a real-valued $\beta_1$ unit change in the outcome variable, in logistic regression models, a one-unit change in the value of the predictor results in a logit probability increase of $\beta_1$.

For the dative data, we have the following interpretation. The intercept says that when $x = 0$, the logit probability estimate is -2.34116 units. The parameter estimate for the length of the recipient says that when the length of the recipient increases by 1 unit (here, one word), the logit probability of getting a PP outcome increases by 0.70085 logits.

Unlike probabilities, the logits range from negative infinity to positive infinity. A logit of 0 indicates a probability of 0.5, i.e. there is no effect. Negative scores indicate probabilities between 0 and 0.5, and positive scores indicate probabilities between 0.5 and 1. Thus, the closer to zero the logit is, the less of an impact there is, while the closer the score tends to infinity, the greater the impact.

However, logits are very difficult to interpret straightforwardly. It is useful therefore to make use of the odds ratio by exponentiating the coefficient (that is, by taking its antilogarithm). Odds are very
CHAPTER 7. METHODOLOGICAL PRELIMINARIES

straightforward to interpret, and are multiplicative. Exponentiating the coefficient for the intercept, it means that when the length of the recipient is 0 (ignore the ridiculousness of this at present), the odds of a PP-outcome is \( \exp(-2.34116) = 0.09621596 \), which means that it is much less likely than a NP-outcome. Exponentiating the coefficient for length of the recipient, it means that when the length of the recipient increases by one unit (i.e. one word), the odds of a PP-outcome increase by a factor of \( \exp(0.70085) = 2.015465 \).

Again, the standard errors indicate how much variability there is with respect to the estimate, and they can be used to construct confidence intervals. For the sample size here, the confidence intervals are equal to \( \beta \pm 1.96 \times SE \). For example, for the intercept, one would expect in 95% of samples, the coefficient to lie between \(-2.34116 \pm (1.96 \times 0.07951)\) in logits, for example. If the confidence intervals include 0 (in logits) or 1 (in odds), then there is no effect of the predictor, because the population intercept could well be 0 (in logits) or 1 (in odds), indicating no effect. For the slope, one would expect in 95% of samples, the coefficient to lie between \( \beta \pm (1.96 \times 0.03866) \). Incidentally, both of these values do not include 0 in logits, so we can conclude that the estimates computed from the sample are reliable and reflective of the population estimate.

Then, just as a t-value is computed by dividing the estimate by the standard error, here we do the same, to get a z-value (some statistics packages report the Wald statistic, which is the squared estimate divided by the standard error, and has a chi-square distribution). The z-value has a standard normal distribution, and compared to critical value for this number of degrees of freedom, it is a significant parameter estimate. Some handbooks recommend avoiding using the z-score as an indicator of parameter significance, because in some samples inflated standard errors may result in underestimated z-values, and hence increase the likelihood of committing a Type II statistical error (not accepting the alternative hypothesis that \( \beta = 0 \), when it should be accepted). Consequently, it may be that a model is significantly better at improving the fit of the model via deviance statistics, but it does not demonstrate an influential p-value.

7.5.3 Multilevel generalized linear models

7.5.3.1 Introduction

A crucial assumption about data in OLS and standard GLMs is that the errors are independent of each other. Linguistic data often violate these assumptions. We sample many and different numbers of observations from the same speaker using the same and different words (“items”), meaning that the data has not been independently sampled. This is probably not of issue if this structure has no effect on the outcome variable, but it is if it is left out of the model specification.

In this section, as a demonstration of GLMERS for the analysis to follow in subsequent chapters, we will continue working with the Bresnan dative dataset, this time restricting the modality to spoken corpora, so that we can explore random variability with respect to the speaker identity (“by-speaker” effects) and verb lemma effects (“by-verb”) effects. This dataset is hierarchically structured, with different numbers of observations taken from different speakers and different numbers of observations taken from verbs. For instance, speaker S1104 has 40 observations, while speaker S1699 has a single observation 1. Similarly, the verb give has 1263 observations recorded on it, while award only has one. There are several problems here: first, there is a statistical reason, as already alluded to above, in that the observations are not independently sampled, and, second, as the verb give dominates, this would not be a model about the dative alternation, but about the verb give, especially if variation within these random components significantly impacts on the response variable. Consequently, it looks like we may have to take this variation into account.

78
7.5.3.2 Variance components models

We start with variance components models, which are sometimes termed random effects models. We saw earlier with GLMs that there is a null model with just the intercept. The variance components model is like the GLM equivalent of a null model. It contains just the random effect levels, and is the baseline to which we compare subsequent models.

It is useful to determine whether a variance components model is in fact needed as the baseline, or whether we could just get away with the null model of the GLM. The reason for this is that adding multilevel structure to the model increases its complexity. In the parlance of Occam’s Razor, we want a model to be parsimonious, in that it contains everything we need it to contain but nothing else.

We (i) first construct a null model based on a GLM, that is a model fitted with just the intercept ($\sim 1$), (ii) a variance components model with the random effect of verb, and (iii) a third model with random effect of speaker. We compare the models in terms of the $-2LL$, that is their deviances, and in terms of their AICs. The model that exhibits the greatest significant deviance from the null model (or the model used for comparison), as measured on a $\chi^2$-distribution with $df_j - df_{null}$. We see from Table 7.4 that the vcm with speaker information in it is not statistically significantly more informative than the null glm, 2440.1 - 2439 = 1.1 and its AIC is higher. By contrast, the vcm with verb in it is more informative (2440.1 - 2002 = 438.1, $df = 2 - 1$, $p < 0.001$), and its AIC is much lower. To conclude, verb bias is an influential random component in the model, while speaker variation is not. As such, we have evidence that the model containing verb is of importance, exerting influential hierarchical/multilevel structure on the response outcome, and we should not consider further models without it. By contrast, the effect of speaker is so weak as to be irrelevant. Consequently, we can disregard this type of variation from further analysis.

<table>
<thead>
<tr>
<th>Model</th>
<th>$-2LL$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM (Null)</td>
<td>2440.1</td>
<td>2442.1</td>
</tr>
<tr>
<td>vcm (Verb)</td>
<td>2002</td>
<td>2006</td>
</tr>
<tr>
<td>vcm (Speaker)</td>
<td>2439</td>
<td>2443</td>
</tr>
</tbody>
</table>

Table 7.4: Comparison of a null GLM with various vcms for the Bresnan et al. (2007) dataset

7.5.3.3 Random intercepts and random slope models with a categorical covariate

In this section we examine the addition of a predictor and random slopes. Let a categorical covariate $x$ have $k \geq 2$ levels. As with classical linear regression, we introduce a categorical covariate into a multilevel model as $k - 1$ dummy variables, coded in whichever way suits. As discussed above, the method we utilize in this study is treatment coding, whereby (0-1)-coding is used: if an observation has a particular attribute, it is coded 1, and if it does not it is coded 0. The baseline level is coded 0 throughout, in order to prevent perfect collinearity. The example categorical predictor here is the semantics of the verb, and it has five levels. Four levels are entered into the regression analysis, with respect to the following design matrix:

```r
> contrasts(dative.new$SemanticClass)
   c f p t
a 0 0 0 0
f 0 1 0 0
p 0 0 1 0
t 0 0 0 1
```

This says that level a is the baseline level (as it coded 0 throughout). We could change the baseline category, but the level a has the most number of observations for comparison, so we keep this because of
the greater statistical power that results.

A variance components model, outlined above, has already indicated that we need to include by-verb variability in the model. Thus, we proceed directly to the random intercepts model, which is the variance components model plus the terms for the predictor variable (4 parameters are being estimated with respect to this predictor). This model is termed dative.semRI. An ANOVA of the VCM and this model suggests that the addition of semantic class is of importance.

\[
\begin{array}{cccccc}
\text{Df} & \text{AIC} & \text{BIC} & \text{logLik} & \text{Chisq} & \text{Chi Df} & \text{Pr(>Chisq)} \\
\text{dative.vcm} & 2 & 2005.7 & 2017.2 & -1000.86 & & \\
\text{dative.semRI} & 6 & 1968.3 & 2002.9 & -978.17 & 45.369 & 4 \times 3.332e-09 \ *
\end{array}
\]

One would then go on to include random (or varying) slopes. While the inclusion of random intercepts assumes that only intercepts vary with respect to a speaker, item, etc., this is not assumed with random slopes. Random slopes specifically take into account the differing behaviours of speakers, items, etc., with respect to how a predictor responds to the outcome variable. In a sense, it is like an interaction. We configure a model for the dative choice as a function of the semantic class predictor, a varying intercept for verb, and a varying slope for verb. This model is termed dative.semRS. It is then compared with the nested model, in this case the model without the random slopes (dative.semRI), in terms of its deviance and AIC. As the R output shows, the inclusion of random slopes is significantly enhancing: semantic class has influence on the choice of construction, but it differs depending on the verb in question.

\[
\begin{array}{cccccc}
\text{Df} & \text{AIC} & \text{BIC} & \text{logLik} & \text{Chisq} & \text{Chi Df} & \text{Pr(>Chisq)} \\
\text{dative.semRI} & 6 & 1968.3 & 2002.9 & -978.17 & & \\
\text{dative.semRS} & 20 & 1784.2 & 1899.5 & -872.11 & 212.13 & 14 < 2.2e-16 \ *
\end{array}
\]

Once this process has been completed, one can then inspect the parameters of the model, and interpret them in the usual way.

7.6 Explanation of predictor activity

When the statistical models have been reported and interpreted, we will turn to an omnibus discussion and, usually, attempt to explain why a predictor acts in the way it does in general linguistic terms. Statistical analysis can only tell us what is important, not why it might be important (association does not indicate cause). In some cases, it might be self-evident, given the previous literature. The explanations here will typically be sought in three broad arenas: generative feature-checking theory (Chomsky, 1995; see Adger, 2003 for an overview); complexity theory; and optimization theory. These concepts are better understood with reference to the predictors themselves, so nothing more will be said about them here.

7.7 Summing up

This chapter has laid the methodological foundations for the subsequent chapters that span the PAC’s information space in detail. The familiar case study of the dative alternation of Bresnan et al. (2007) in the languageR package (Baayen, 2013) has provided a sufficient background of the quantitative techniques, serving as a constant reference point for the chapters which follow.
Chapter 8

Multilevel Structure

8.1 Introduction

In the previous chapter I discussed statistical problems with independence and how higher level “random effect” structures (such as, for instance, by-item and by-speaker differences) can have an important influence on variability. For example, different speakers may show a preference for different variants, and the effect that a predictor variable has on the response variable may differ by speaker. The former is captured by random (or varying) intercepts, the latter by random (or varying) slopes. Accordingly, this chapter motivates the presence of higher level structure in the information space of PAC modeling.

There are two possible loci of higher level effects in the present dataset, viz. the text and the verb’s lemma. As will be discussed below, this is because we have sampled different numbers of observations from different texts, and likewise we have sampled different numbers of observations from different verbs. This is characterized in a somewhat simplified fashion in Figure 8.1, for text.

This chapter is structured as follows. We begin in Section 8.2 by presenting a null model of a standard glm with just the intercept fitted. In Section 8.3 we consider introducing higher level variance of text, and in Section 8.4 lemma-based higher level variance is explored. The models considered in those latter two sections are single multilevel models with just one higher level variable present. In Section 8.5, however, we consider whether a more complex multilevel model is of statistical benefit. Finally, in Section 8.6 a summary is presented, illustrating how we will use this model in the subsequent chapters.

8.2 The glm null model

In order to motivate the presence of higher level features, such as text and lemmatic variation, into the modelling process it will first be necessary to construct a null generalized linear model. This model has

Figure 8.1: Multilevel by-text structure characterizing the PAC dataset. The individual observations are referred to as i-units, whereas the higher levels are referred to as j-clusters. The i-units make up the clusters j, and the j-clusters make up the total sample n. The same characterization could be done for verb lemma.
the following structural form:

$$\text{logit}[P(y = 1)] = \beta_0 + \epsilon$$  \hspace{1cm} (8.1)

This model asserts that the logit probability of the response outcome being equal to 1 (where 1 is to be understood as referring to the PAC serialization *factus est*) is a linear function of only the intercept parameter $\beta_0$ and an error term $\epsilon$. I refer to the object created out of this model as glm.null. It will be of essential importance in the next two sections, but we will have no reason for referring to it in further chapters, unless of course we find that text-based and lemma-based variability exerts no influence on the response variable. The model is constructed in the freely available statistics package R (R Development Core Team, 2012), as with all the statistical models discussed in this thesis. When input into R using the present dataset, Equation 8.1 has the following form in the R programming language:

```r
> glm.null <- glm(Serialization~1, data=pac, family=binomial)
```

We will now inspect the requisite model information for this object, namely the deviance ($-2LL$) and the Akaike Information Criterion ($AIC$), as well as some other interesting pieces of information.

```r
> glm.null$null.deviance
[1] 3238.803
> glm.null$aic
[1] 3240.803
> glm.null$df.null
[1] 2408
> exp(0.41343)
[1] 1.511995
```

We can see that the null deviance for glm.null is 3238.803, and its $AIC$ is 3240.803. Further, there are 2408 degrees of freedom, since there are 2409 observations and one parameter (the intercept) is being estimated (recall that the formula for the degrees of freedom for a model is calculated as $n - p$, where $n$ is the number of observations and $p$ is the number of parameters being estimated). In what follows, we will compare these values to those obtained for the subsequent models in this chapter. Finally, it is worth observing that the intercept for the null model is 0.41343, which means that the serialization *factus est* is 1.511995 times more likely to occur than *est factus* (which equates to $\text{logit}^{-1}$ of 0.60191, meaning that *factus est* occurs on average roughly 60% of the time). We already knew this information from the univariate frequency summary in 6.4, but it worth spelling it out in terms of statistical modelling.

### 8.3 Text-based variability

We have already seen from our sampling summary in Table 6.3 that there are fluctuations between texts in the choice of PAC serialization (cf. Table 8.1, which summarizes this information). Furthermore, different numbers of observations were taken from each text, violating the assumption of independence (which assumes that exactly 1 observation was randomly sampled from each text). It therefore seems appropriate to consider text-based information as a higher level random effect, and potentially allow mainstream predictor variables to interact with it. Here, we examine whether the null model fares significantly worse than one that includes the random effect of text using deviance and $AIC$ statistics. As discussed earlier, this model is termed a *variance components model*. It will be referred to as glmer.vcm.text. If the model including higher level text-based information exhibits a significant degree of deviance between it and the null model and results in a lower $AIC$, then such information is clearly exerting an important influence on the probability of the outcome variable, and it should not be excluded. If, by contrast, the deviance between it and the null model is not significant, it can be disregarded from further consideration.
The model of interest has the following functional form:

\[
\text{logit}\left[P(y_{ij} = 1)\right] = \beta_0 + u_j + \epsilon_{ij} \tag{8.2}
\]

Here, \(j\) is the text, \(\beta_0\) is the overall grand mean of \(\text{logit}\left[P(y)\right]\) across all texts, \(\beta_0 + u_j\) is the mean of \(\text{logit}\left[P(y)\right]\) for the \(j\)th text, and \(\epsilon_{ij}\) is the difference between the \(y\)-value for the \(i\)th observation and that observation’s text mean. In \(\text{R}\), we have the following code for this particular equation:

```r
glmer.vcm.text <- glmer(Serialization~(1|Text),data=pac,family=binomial)
```

From this model, we extract the relevant statistics:

Generalized linear mixed model fit by maximum likelihood ['glmerMod']
Family: binomial (logit)
Formula: Serialization ~ (1 | Text)
Data: pac

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>3149.609</td>
<td>3161.183</td>
<td>-1572.805</td>
<td>3145.609</td>
</tr>
</tbody>
</table>

Random effects:
Groups Name Std.Dev.
Text (Intercept) 0.417

Number of obs: 2409, groups: Text, 9

Fixed Effects:
(Intercept) 0.2027

Calculating the significance of the deviance difference between `glm.null` and `glmer.vcm.text`

\[93.194 = 3238.803 - 3145.609\] on 2-1 degrees of freedom, we find it to be a highly significant difference (\(p < 0.001\)). We also note that the \(\text{AIC}\) of `glmer.vcm.text` is lower than that of `glm.null` (it is closer to zero), indicating that the model benefits from including the higher level term of text. Finally, the intercept says that the grand mean of `factus est` across the texts is \(\exp(0.2027) = 1.224705\) more likely than `est factus`. This is clearly different from the above, indicating it occurs in 55.1% of cases. This is very close to the grand mean calculated by hand (= 54.2223).

It is also important to think in terms of adjustments to the intercept for each text. These are displayed in Table 8.2 and plotted in Figure 8.2. The adjustments to the intercept, stored as `ranef(model)` in \(\text{R}\) indicate by how much the intercept moves up or down for different texts. To give an example, the
The grand mean intercept is 0.2027, but in the text B.Afr. it has to be adjusted down by $-0.42919096 = 0.2027 - 0.42919096 = -0.226491$. This indicates that in the B.Afr. the serialization *factus est* occurs 0.7973265 times for every 1 time that it does not occur. As we can clearly see from the graph, the text with the least adjustment required is B.Alex, while the text with the largest adjustment required is Cicero’s Phil.

<table>
<thead>
<tr>
<th>Text</th>
<th>Intercept Adjustments</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. Afr.</td>
<td>-0.42919096</td>
</tr>
<tr>
<td>B. Alex.</td>
<td>0.04387454</td>
</tr>
<tr>
<td>B. Hisp.</td>
<td>-0.34518223</td>
</tr>
<tr>
<td>Caesar Civ.</td>
<td>-0.23608533</td>
</tr>
<tr>
<td>Caesar Gal.</td>
<td>0.38856473</td>
</tr>
<tr>
<td>Cicero Phil.</td>
<td>0.83806759</td>
</tr>
<tr>
<td>Hirtius</td>
<td>-0.19544682</td>
</tr>
<tr>
<td>Nepos</td>
<td>0.20445415</td>
</tr>
<tr>
<td>Varro R.R.</td>
<td>-0.27210864</td>
</tr>
</tbody>
</table>

Table 8.2: Intercept adjustments to a null model for text

In sum, it makes both conceptual and statistical sense to unpool the dataset into its various textual components.

### 8.4 Lemma-based variability

In the previous section we motivated the presence of a random higher level variable of text in the statistical modelling. However, we also have higher level structure with respect to verb lemmas. PAC observations were drawn from 604 verbs, but it may be the case that there is lemma-based variability with respect to the verb that needs taking into account. Lemma-based differences may be less obvious than differences according to text, so we begin the discussion of why lexical effects can be important in language, before
examining its structural effects on the pac alternation.

8.4.1 Background

In traditional structuralist and generative frameworks, lexis has no place in grammar (e.g. Chomsky, 1957). A clear distinction is made between the Grammar, where syntactic computations take place, and the Lexicon where lexis is stored. However, recent usage based approaches, such as Pattern Grammar (Hunston & Francis, 2000) and Cognitive and Construction Grammar (Stefanowitsch & Gries, 2007), have stressed the importance of lexis in grammar (see Romero 2009 for an overview). There is strong psycholinguistic evidence, both in production and in processing, for the importance of lexis in grammar. In addition, recent research in corpus linguistics has shown that lexis is prominent in many types of grammatical alternation, with certain lexemes strongly preferring one construction to (an)other possible structural variant(s). For instance, when the choice is between an active and passive construction, the lexeme base opts for the passive in a probabilistic, non-deterministic way. We shall term the interaction between lexis and syntax, the lexis–syntax interface, and the extent to which lexemes exhibit preferences for different constructional variants lexical bias.

8.4.2 Reported lexical effects on grammatical variability

8.4.2.1 General literature

There is a good deal of corpus evidence for the impact of lexis both in univariate and multivariate terms. For instance, Wasow (2002: 87–88) and Wasow & Arnold (2003: 132–134) briefly report on lexical effects in the English dative alternation (to-dative vs. double object). They examine between 50 and 100 observations each from five verb lemmas participating in this alternation — namely, give, hand, bring, send, and sell — sampled from the New York Times. They find that, for example, the verb give favours the double object construction, while the verb sell favours the to-dative construction.

Additionally, Gries & Stefanowitsch (2004) present a statistical methodology to explore the interaction between lexemes and constructional variants, termed “distinctive collexeme analysis”. This method seeks to test whether a certain “collexeme”, or lemma λ₁, is statistically significantly “distinctive” in one constructional variant α versus another β, by computing a Fisher exact test over a matrix that contains corpus frequencies for (i) the lemma λ₁ in variant α, (ii) the lemma λ₁ in variant β, (iii) the set of all lemmas Λ in variant α − λ₁ in variant α, and (iv) the set of all lemmas Λ in variant β − λ₁ in variant β. If the resulting p-value is below < 0.05, or in later research above − log₁₀(p) = 1.30103, then that lemma is distinctive in that variant. This procedure is then iterated for the remaining lemmas (λ₂...λₙ) in Λ that participate in the variants (α, β). To exemplify their method, a number of case studies are presented — dative, active/passive, verb–particle, will/be going to, and genitive alternations. These case studies testify to the robustness of lexis on grammatical alternatives. To exemplify, in the dative alternation, give, tell, show, offer, and cost are distinctive in the ditransitive (= double object) construction, whereas bring, play, take, pass, and make are distinctive in the alternative to-dative construction, results which overlap with the findings of Wasow (2002) and Wasow & Arnold (2003), mentioned above.

8.4.2.2 AVC literature

If we zoom in on the AVC literature specifically, we also find that lexical factors have been observed for serialization differences with respect to these constructional types in, e.g., Dutch and Latin.

With respect to Dutch, De Sutter (2005), exploiting the methodology of Gries & Stefanowitsch (2004), finds that the top five participial lexemes drawn to Dutch participle_AUXILIARY order are bedoeld, bedreigd, begrepen, bekend, and beperkt, while aangenomen, gebracht, gegaan, gegeven, and gegoooid are drawn to the auxiliary-participle order (2005: 246). The differences are interpreted in constructional semantic terms, rather than lexically, however. It is suggested, following the literature, that participle lexemes
preferring the participle-auxiliary order profile states, whereas those participle lexemes preferring the opposite profile typically profile actions.

Turning to the Latin pac, Marouzeau (1910: 116) remarks that certain verbs have a “sorte de prédetermination” to occur in particular pac variants because of their inherent semantics. The example Marouzeau cites is the verb audere ‘to dare, to have the audacity to’ occurring in est factus and est...factus variants. He relates this tendency to his explanation that est factus and est...factus are used in strongly affirmational contexts (1910: 114f. — see later).

Devine & Stephens (2006: 189–190) observe serialization differences with respect to individual verbs. Specifically, there is “a striking contrast”, noticed in ut-clauses in Caesar, between (i) the lemmas demonstro ‘to demonstrate’ and dico ‘to say’ and (ii) the lemmas impero ‘to command’ and praecepio ‘to order’, in that the former pair favours the factus est serialization with “multiple occurrences” and the latter pair “regularly” favours est factus serialization. These differences are argued to correlate with semantic differences (eventive vs. stative), to be discussed later, but given that semantic effects are nuanced and subtle, it may be lexis that is driving the variation here. “[ut demonstratum est] compares the event described by the clause with a previous command while [ut erat imperatum] says that the discourse contains an earlier speech act making the same assertion” (2006: 190). However, this alleged connection strikes me as being rather tenuous and hard to pin down.

8.4.3 Annotation

Against this backdrop, I will explore the possibility that the lexical details of the verb’s lemma might account for some of the variability between factus est and est factus. The lemma is coded as listed in Oxford Latin Dictionary (Glare, 1982), and termed LemmaParticiple. For example, missus est ‘he has been/was sent’ and missi sumus ‘we have been/were sent’ are tagged for their lemma mitto. Some points are perhaps worth noting. First, for participles that are adjectives or may be used as adjectives, I code — for the sake of simplicity — the verb from which the participle is ultimately derived: thus, tutus ‘safe’ is coded under the same lemma as tuitus ‘watched over’, tueor. Second, I treat factus ‘made, done’ from facio and factus ‘became, happened’ from fio as facio, again for simplicity but also because clauses involving them can be semantically ambiguous, especially in subject complement structures: Tarquinius rex factus est = ‘Tarquinius was made king’ or ‘Tarquinius became king’.

8.4.4 Multilevel modelling

With the above background in place, we now turn to an examination of whether the null model fares significantly worse than one that includes the random effect of verb lemma using deviance and AIC statistics, just as we did for textual information. It will be referred to as glmer.vcm.lemma. If the model including higher level lemma-based information exhibits a significant degree of deviance between it and the null model and results in a lower AIC, then such information is clearly exerting an important influence on the probability of the outcome variable, and it should not be excluded. If, by contrast, the deviance between it and the null model is not significant, it can be disregarded from further consideration.

The model of interest has the following functional form:

\[
\text{logit}\{P(y_{ij} = 1)\} = \beta_0 + u_j + \epsilon_{ij}
\]  

(8.3)

Here, \(j\) is the text, \(\beta_0\) is the overall grand mean of \(\logit[P(y)]\) across all verb lemmas, \(\beta_0 + u_j\) is the mean of \(\logit[P(y)]\) for the \(j\)th verb lemma, and \(\epsilon_{ij}\) is the difference between the \(y\)-value for the \(i\)th observation and that observation’s verb mean. In \(R\), we have the following code for this particular equation:

```
glmer.vcm.lemma <- glmer(Serialization~(1|ParticipleLemma), data=pac, family=binomial)
```

\(^1\text{Cf. Marouzeau (1910: 52): “Certains mots sonts pour ainsi dire prédestinés par leur sens même occuper une place déterminée dans leur groupe”}.

86
8.4. LEMMA-BASED VARIABILITY

From this model, we extract the relevant statistics:

**Generalized linear mixed model fit by maximum likelihood ['glmerMod']**

- **Family:** binomial (logit)
- **Formula:** Serialization ~ (1 | LemmaParticiple)
- **Data:** pac

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3230.001</td>
<td>3241.575</td>
<td>-1613.001</td>
<td>3226.001</td>
</tr>
</tbody>
</table>

**Random effects:**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LemmaParticiple</td>
<td>(Intercept)</td>
<td>0.4679</td>
</tr>
</tbody>
</table>

**Number of obs:** 2409, **groups:** LemmaParticiple, 604

**Fixed Effects:**

- (Intercept) 0.401

```r
> attr(logLik(glm.vcm.lemma), "df")
[1] 2
> exp(0.401)
[1] 1.493317
```

Calculating the significance of the deviance difference between `glm.null` and `glm.vcm.lemma` 12.802 = 3238.803 − 3226.001 on 2-1 degrees of freedom, we find it to be a highly significant difference ($p < 0.001$). We also note that the **AIC** of `glm.vcm.lemma` is lower than that of `glm.null` as it is closer to zero, indicating that the model benefits from including the higher level term of verb lemma information. Finally, the intercept says that the grand mean of *factus est* across the verb lemmas is $\exp(0.401) = 1.493317$ times more likely than *est factus*.

It is also important to think in terms of adjustments to the intercept for each verb lemma. All these are plotted in Figure 8.3. Further, a selection of intercept adjustments are reported in Table 8.3, which displays the top fifteen verbs which strongly associate with *est factus*, and thus result in negative adjustment values, and the top fifteen verbs that strongly associate with *factus est*. The adjustments to the intercept, stored as `ranef(model)` in R indicate by how much the intercept moves up or down for different verb lemmas. To give an example, the grand mean intercept is 0.401 (in logits), but for the verb *contineo* ‘contain’ it has to be adjusted down by $-0.688402059$: 0.40096 − 0.688402059 = $-0.2874421$. This indicates that for *contineo* the serialization *factus est* occurs 0.75018 times for every 1 time that it does not occur. As we can clearly see from the graph, the verb lemma with the greatest adjustment required is *demonstro* ‘point out’, which is almost categorically realized as *factus est* (94.44%). One possible reason that this is so will be explored in chapters on the pac’s semantics and morphological structure. The verb *contineo* ‘contain’ is strongly stative and is usually used as an adjective in the sense ‘content, happy’ in the present dataset rather than as a verbal participle. By contrast, the verb *demonstro* ‘point out’ is strongly eventive, punctual, and verbal. It may be this distinction that results in different word order preferences for these verb lemmas. However, one should note that, as pointed out above, lemma distinctions may be in and of themselves important, separate from their semantic construal. We can also relate the present findings to previous research. According to the analysis here, it does not appear Marouzeau’s (1910) claim extends to this dataset, as *audeo* ‘dare’ prefers associating with *factus est* rather than *est factus*, as he claimed. Devine & Stephens (2006) appear to make some correct observations, however, in that *impero* ‘enjoin upon’ and *demonstro* ‘point out’ preferentially select *est factus* and *factus est*, respectively. While *praecipio* and *dico* do not feature in the top 30 verbs in the table, the former exhibits an adjustment to the intercept in the direction claimed by Devine & Stephens (2006), namely 0.34, but *dico* hovers just below the intercept, namely −0.01. Consequently, *dico* is a “well-behaved” verb, showing little in the way of lexical bias.
CHAPTER 8. MULTILEVEL STRUCTURE

Verb Lemma | Translation | Intercept Adjustment | \( n_j \) \\
--- | --- | --- | ---
continuo | contain | -0.69 | 19
instruo | draw up | -0.59 | 6
edo.2 | publish | -0.52 | 5
ago | drive | -0.51 | 7
perfero | carry through to | -0.51 | 7
scribo | write | -0.49 | 11
dedo | surrender | -0.43 | 4
deligo.2 | tie up | -0.43 | 4
includo | shut up | -0.43 | 4
expello | banish | -0.43 | 6
perficio | complete | -0.42 | 10
impero | enjoin upon | -0.42 | 14
affero | carry to | -0.36 | 13
accipio | receive | -0.35 | 7
obicio | throw at | -0.35 | 7
paro | prepare | 0.33 | 17
hortor | encourage | 0.35 | 5
laetor | rejoice | 0.35 | 5
polliceor | promise | 0.38 | 15
arbitror | think | 0.41 | 6
sero | sow | 0.41 | 6
nascor | be born | 0.41 | 29
imemio | find | 0.42 | 16
audeo | dare | 0.43 | 33
loquor | speak | 0.44 | 10
cognosco | find out | 0.44 | 20
conor | try | 0.45 | 17
coeipi | begin | 0.51 | 15
patior | allow | 0.60 | 14
demonstro | point out | 0.74 | 18

Table 8.3: Intercept adjustments to a null model for verb lemma

8.5 Complex multilevel structure

One might ask whether both by-text and by-lemma information is needed. We assess this again by using \(-2LL\) and \(AIC\) statistics, using an analysis of deviance ANOVA between a complex multilevel variance components model that includes information on both text and lemma and two nested models, one containing information on just text and the other containing information on just lemma. The latter two are equivalent to the models that have just been presented, namely \texttt{glmer.vcm.text} and \texttt{glmer.vcm.lemma}, respectively. We will call the former more complex model \texttt{glmer.vcm.both}.

Tables 8.4–8.5 are two ANOVA tables. In the first table, we have a comparison of the deviances for \texttt{glmer.vcm.text} and \texttt{glmer.vcm.both}. This says that the model including both higher level variables, text and verb lemma, is a significantly better than a nested model that just includes by-text information. In the second table, we have a comparison of deviances for \texttt{glmer.vcm.lemma} and \texttt{glmer.vcm.both}. Similarly, this says that including both higher level variables, text and verb lemma, is a significantly better model than a nested model that just includes by-lemma information. The fuller model thus significantly reduces the deviance than either of the two nested models.
"implausibility" in most scenarios, cf. Gelman & Hill (2007: 283). In argument (where
\(j\) is the predictor and \(x\) is the group level variable), are not considered in this study because of their
"implausibility" in most scenarios, cf. Gelman & Hill (2007: 283). In \(\hat{\eta}\), if varying slopes are configured, varying intercepts
are configured as a default.

8.6 Summary

In this chapter we have examined two potential sources of higher level variation with respect to the
response variable, the choice between \textit{factus est} and \textit{est factus}. We have shown that a minimally adequate
baseline model with random structure contains both by-text and by-lemma information. In the following
chapters, this baseline model will be used to assess the relevance of each predictor in turn. For each
predictor variable discussed in the following chapters, three models will be considered: (i) the baseline
variance components model as constructed in this chapter,\(^2\) (ii) the random intercept model with the
higher level structure and the predictor in it, and (iii) the random slope model. The model that will be
reported in full and interpreted (for each predictor) is the model that is the optimal in statistical terms
(i.e. in its deviance and AIC statistics).\(^3\)

\(^2\)Actually, what we have presented here is the baseline variance components model with all of the observations included.
Some of the predictor variables have missing information on them. Consequently, for these models, separate baseline
variance components models are constructed.

\(^3\)As discussed earlier, varying slope models without varying intercepts, while possible to configure in \(\hat{\eta}\) with a \((x\mid j)\)
argument (where \(x\) is the predictor and \(j\) is the group level variable), are not considered in this study because of their
"implausibility" in most scenarios, cf. Gelman & Hill (2007: 283). In \(\hat{\eta}\), if varying slopes are configured, varying intercepts
are configured as a default.

Table 8.4: \textsc{anova} of \texttt{glmer.vcm.both} and \texttt{glmer.vcm.text}

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{glmer.vcm.text}</td>
<td>2</td>
<td>3149.61</td>
<td>3161.18</td>
<td>-1572.80</td>
<td>3145.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\texttt{glmer.vcm.both}</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td>12.20</td>
<td>1</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

Figure 8.3: Intercept adjustments to a null model for verb lemma
### Table 8.5: ANOVA of `glmer.vcm.both` and `glmer.vcm.lemma`

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>glmer.vcm.lemma</td>
<td>2</td>
<td>3230.00</td>
<td>3241.58</td>
<td>-1613.00</td>
<td>3226.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>glmer.vcm.both</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td>92.59</td>
<td>1</td>
</tr>
</tbody>
</table>
Chapter 9

Phonology

9.1 Introduction

Phonological variables are the subject of this first chapter that discusses the fixed effects in the information space of the PAC alternation. There are several reasons why one might want to consider phonology as being interesting to the present case study, pertaining to how it is thought to relate to syntax.

As is well known, the standard Chomskyan view of language is that it is modular: syntax and phonology are separate modules within the architecture of grammar (Chomsky, 1995). The only way in which they are related is via the syntax-phonology interface, and that interface is claimed to be unidirectional: while the syntax can drive the phonology, the phonology cannot drive the syntax. Zwicky (1969) and Zwicky & Pullum (1986) dub this “Phonology-Free Syntax”. However, this view has been challenged via empirical evidence. For example, Schlüter (2005) adduces a wide range of evidence that English grammar is sensitive to rhythmical features, and Shih & Zuraw (2014) claim that phonology is an influential constraint on adjective-noun serialization in Tagalog, ranging from syntactic obedience to the Obligatory Contour Principle (OCP) to syntactic effects of prosodic structure. Other studies have found that while syntax is sensitive to prosodic phonology, it is less sensitive to segmental phonology in predictable ways (Jaeger, 2006).

Against this theoretical backdrop, the present chapter therefore probes the effect that phonology has on the PAC alternation. Several phonological variables will be considered: clisis and prosodic structure in Section 9.2; two types of phonological optimization, rhythmical optimization in Section 9.3 and segmental optimization in Section 9.4; and clausula in Section 9.5.

9.2 Clisis and Prosodic Structure

We begin by addressing clisis and its impact on serialization choice in the Latin PAC. It is necessary to begin with a definition, and justify at the outset its relevance to the information space considered in this study.

Clitics — such as the English auxiliary ‘s — have a special status in the grammar. In terms of where they belong as a grammatical category, first, they are distinguished from prosodic words in that (i) they are prosodically deficient in their phonology (whether that is in terms of stress, tone, or pitch), and (ii) they require a prosodically full word to their left (if they are enclitic) or to their right (if they are proclitic) to combine with. Second, they are distinguished from morphological affixes in that, while affixes are very selective (for instance, they combine with an entity which they are morphosyntactically related to), clitics, by contrast, exhibit a low degree of selectivity, attaching to whatever element is in their local proximity; Spencer & Luís (2012) dub this characteristic of clitics “promiscuous attachment”.

In the literature two types of clitics have been distinguished — simple clitics which have the same
distribution as their corresponding full variant (e.g. English auxiliary ‘s occurs in the same serial position as is/has in declarative sentences) and special clitics which typically have a different distribution to their corresponding full variant (examples below). The latter type — the special clitics — often occupy the second position (2P) in the clause: depending on the language, this can be second word after a prosodically full word (2W) or second after the first syntactic constituent in the nuclear clause or phonological phrase (2D).

To give an example of a clitic, in the following example from Slovene, the clitic je is unstressed and phonologically adjoined to the NP moje srce that precedes it.

(1) in moje srce je bilo veselo and my heart AUX.3.SG be.LPART happy ‘and my heart was happy’

(Slovene, Spencer & Luís 2012: 54, ex. 33a)

As we shall see below, auxiliaries are one of a number of grammatical categories that are often clitic or have a clitic counterpart in addition to a full variant. It has been demonstrated in Latin, too, there were clitic and non-clitic auxiliaries. Consequently, it seems to be appropriate for us to explore the possible relevance of clisis to the Latin PAC alternation.

9.2.1 Latin auxiliaries and clisis

As the example in (1) suggests, auxiliaries are one of a range of grammatical categories that can phonetically reduce and become clitics, alongside (e.g.) prepositions, complementizers, and determiners (Zwicky, 1977).

This has also been observed to be partly the case in Latin, where there is evidence that at least two forms of the auxiliary esse ‘to be’, namely the second person singular present indicative es and the third person singular present indicative est, have clitic allomorphs, -s and -st, respectively (see e.g. Wackernagel 1892; Allen 1973; Wanner 1987; Adams 1994; Devine & Stephens 2006; Fortson 2008; Weiss 2009; Pezzini 2011, 2012). (In the discussion that follows I will use est to refer to the phonetically full forms es and est, -st to refer to the clitic forms -s and -st, and (e)st when it is unclear.)

Pezzini (2011, 2012) has recently marshalled a wide range of evidence, ranging from inscriptive data, the language of the early comedians such as Plautus and Terence, languages geographically local to Latin such as Oscan and Umbrian, to the testimonies of ancient grammarians, in support of this. To give an example, in (2a) the auxiliary -st is a clitic form, and is incorporated with the participle natus ‘born’ on its left edge, with concomitant simplification of the resulting consonant cluster natuss → natust. This can be compared with the non-clitic variant, given in (2b).

(2) a. puer natu=st
   boy.M.SG.NOM born.PTCP.PRF.M.SG.NOM=be.3.SG.PRS.IND
   ‘A boy has been born’ (Ter. Ad. 728)

b. natus est nobis nepos
   born.PTCP.PRF.M.SG.NOM be.3.SG.PRS.IND us.PRN.M.PL.DAT grandson.M.SG.NOM
   ‘A grandson has been born to us’ (Ter. Hec. 639)

1The distinction between simple and special clitics goes back to the classical paper by Zwicky (1977), but there is a different opinion, and its usefulness has recently been challenged by Bermúdez-Otero & Payne (2011) who argue that that position is untenable.

2Note, however, that according to Pezzini (2011) “the evidence for contraction of the second person form es is considerably inferior to, and qualitatively different from, that of est”.

3There is no explicit evidence that other forms of esse ‘to be’ can be clitic. We don’t find, say, -rat for erat. However, this does not necessarily mean that they might not have been phonologically reduced or contracted in some way (see e.g. Adams (1994) citing Lindsay (1894) on the third person singular imperfect indicate form erat and its continuation into Italian as era – with an open e rather than the expected ie diphthong, a possible indication of its phonetically reduced status). A similar point can be made for English was, in which the vowel can be reduced to a schwa [w@z], but it does not have a contracted clitic alternate. I return to this point below.
According to Pezzini (2012: 169), in the early Latin of the comedian Terence, the clitic form -st is regular in PACS that require a third person singular present form of esse: the auxiliary is realized as -st in 31 instances (64.58%), est in 17 instances (35.42%), which compares with a combined rate for the verb (i.e. in auxiliary, copula, existential, and locational function) of 54 instances (39.42%) for clitic -st and 83 instances (60.58%) for non-clitic est.\(^4\) In contrast, there is little orthographic evidence for -st in Late Republican Latin prose (Pezzini, 2012: 108).\(^5\) For instance, when there is a choice between est and -st in prose inscriptions, the clitic variant occurs only 1% of the time, and there are “only isolated occurrences in the oldest manuscripts of Cicero” where one might otherwise expect them (Pezzini, 2012: 145). Thus, Pezzini (2012: 145) concludes, “[c]ontracted spellings were therefore apparently avoided for the most part in classical prose”. Be that as it may, as Pezzini notes in various places, absence of the spelling -st cannot be taken as evidence for the absolute absence of a phonetically reduced [st] in the Late Republican period:\(^6\) either (i) prose authors may simply have chosen to reflect, e.g., both [faktus est] and [faktost] as factus est in all environments rather than factus est and factust respectively, or (ii) scribes, who may have been unfamiliar with -st (Pezzini, 2012: 147), might have simply replaced it with est in the manuscripts.

Given this background, we must consequently acknowledge the possibility that some instances of est (and es) in our dataset might in fact reflect what were ultimately phonetically clitic forms, while exercising caution of course, and we must therefore explore how it affects serialization choice. Before we do this, we will in the next section first spell out how clisis has been demonstrated to be relevant to word order generally and in AVCS specifically.

9.2.2 Clitichood and word order

9.2.2.1 General discussion

Clisis is well known to have an influence on word order variability, with, for instance, clitic pronouns typically being in a complementary distribution with phonologically full NPs and contrastive pronouns. For example, in particle verb constructions in English, unstressed/clitic pronouns are not permitted in the continuous construction: compare e.g. Max put out the rubbish vs. *Max put out it (Kim, 1996). Similarly, in dative syntax, while John sent his son the book is grammatical, ?John sent his son it is ungrammatical according to Kim (1996).\(^7\) In several Romance languages, clitic object pronouns appear before the finite verb whereas full NPs occur after it (Spencer & Luis, 2012: 27–28). Contrast the following French examples in (3).

(3)  
   a. Je vois l’homme  
       I see the man  
       ‘I see the man’ (French)  
   b. Je le vois  
       I him see  
       ‘I see him’ (French)

In fact, the syntactic behaviour of clitic pronouns is so peculiar that the WALS database (Dryer & Haspelmath, 2013) excludes constructions involving them from their word order feature catalogue.

\(^4\)Actually, these statistics refer only to (i) PACS with masculine participles, where the metrical evidence for clisis (rather than say, regular elision) is more robust, and (ii) where the alternation is not dictated by metrical constraints. It should be noted that the rate of contraction does not seem to differ that much after feminine and neuter participles, where the auxiliary is spelled as -st in 62% of instances (Pezzini, 2012: 200).

\(^5\)Late Republican Latin verse is a different matter, where -st is more amply attested (Pezzini, 2012).

\(^6\)Pezzini (2012: 150) notes that “it is likely that, as a phonological phenomenon, it was going to disappear (or had already disappeared from speech in the first century BC”, but ultimately concludes that the evidence is largely conflicting.

\(^7\)I mark this example with ‘?’ as I am not convinced as to its complete grammatical unacceptibility, though I do feel it is less grammatical than John sent his son the book.
Turning to morphosyntactic entities other than pronominal clitics, we see that in Czech, whereas
adverbial particles are usually free to move around in the clause, the adverb prý ‘apparently’ is constricted
to occur in 2P (Spencer & Luís, 2012: 19), as the templates in (4) indicate.

(4) a. Eva prý juž napsala ten dopis 
   Eva CL already has.written that letter 
   (Spencer & Luís, 2012: 19) 
   ‘Apparently, Eva has already written that letter’

9.2.2.2 Clisis in AVC research

As noted above, auxiliaries often have clitic (and weak forms) as opposed to full variants. It is thus
no surprise to find that clisis has been mentioned with respect to AVC serialization. To begin with a
straightforward example, in Serbian/Croatian, which has the full auxiliary/copula jesu and the clitic
auxiliary/copula su, only the full auxiliary can occur in first position (1P) in the sentence:

(5) a. jesu 3.pl.cop u sadu 
   3.PL.COP in garden 
   ‘they are in the garden’ (Serbian/Croatian, Spencer & Luís 2012: 16)

By contrast, both jesu and su can also appear in 2P:

(6) a. Devojke jesu u sadu 
   girls 3.PL.COP in garden 
   ‘the girls are in the garden’ (Serbian/Croatian, Spencer & Luís 2012: 16)

In fact the clitic is confined to 2P, such that in (7), where the clitic surfaces in third position — treating
u sadu ‘in the garden’ as a syntactic and phonological phrase — is grammatically illicit.

(7) *Devojke u sadu su 
   girls in garden 3.PL.COP 
   ‘the girls are in the garden’ (Serbian/Croatian, Spencer & Luís 2012: 17)

Given what we have said above in Section 9.2.1 about the potential for two Latin auxiliaries, est and es,
to have clitic variants, -s and -st, to what extent has clisis been argued to impact on serial variability
within the PAC? I note that there is very little by way of discussion here.8

To give an example from my own work, in PACs in which the participle is fronted away from the
auxiliary, the rate of discontinuity between factus est, as in (8a), is low, whereas that between factus
foret, as in (8b), is comparatively high (Brookes 2009). For instance, in the historian Tacitus, in the
factus est serialization, the auxiliary is est (whether that actually ultimately reflects the full form est or
the clitic form -st) in 70% of instances, while it is foret (hardly a clitic auxiliary) in 5% of instances. By
contrast, in the factus . . . est serialization, the auxiliary is est (again, whether that reflects est or -st) in
15% of instances, while it is foret in 40% of instances.

(8) a. profectus retro Tarracomem est 
   depart.PTCP. PRF.M.SG.NOM back Tarraco.ACC be.3.SG.PRS.IND 
   ‘He departed back to Tarraco’ (Liv. 28. 35. 13)

8Adams (1994) argues that esse ‘to be’ in general is a clitic. For him, clisis does have an effect on the position of est in
the sentence, in that while it is clauseally “peripatetic” it typically “[hangs] . . . on the focus of the remark”.

94
b. *admotus* extemplo exercitus *foret*

if move.PTCP.LPF.M.SG.NOM immediately army.M.SG.NOM be.3.SG.IMPF.SBJV

‘if the army had been moved immediately’ (Liv. 43. 4. 1)

I now submit that the reason behind this may be that, while the sequence *factus est*, or rather *factust*, is a prosodic word with clitic attachment of *(e)st*, the sequence *factus foret*, by contrast, consists of two prosodic words and together forms a phonological phrase; this is exemplified in (9). Consequently, if the participle is fronted away from the auxiliary, *foret* can bear its own centre of prosodic prominence, but *(e)st* cannot (assuming there is no suitable host before it).

(9) a. \( \omega \) b. \( \phi \)

\[
\begin{array}{ccc}
\omega & | & \text{Cl} \\
\omega & | & \omega \\
\text{factus} & \text{est} & \text{factus} & \omega & \text{foret}
\end{array}
\]

In a similar vein, Pezzini (2012: 196-170; 179) also offers a number of observations on how differences in the auxiliary’s prosodic status can have an effect on word order in the *PAC*, using data taken from the early comedian Terence. There are distributional differences between clitic -st and non-clitic *est*. For masculine participles, when the auxiliary is *est*, it occurs after the participle in 12 instances (70.59%) and in another position, including directly before the participle, in 5 instances (29.41%); when the auxiliary is -st it can only occur after the participle. A Fisher exact test conducted over Pezzini’s data reveals this to be a statistically highly significant effect (\( p = 0.003614, \phi = 0.3891766 \)). Further, drawing on the information Pezzini (2012: 200, table 31) presents in his table for *(e)st* occurring with feminine and neuter participles (ending in -a and -um, respectively), contraction was substantially preferred when the auxiliary follows the participle than when it occurs in another position (112 for factast vs. 22 instances for -st facta, facta...-st, -st...facta), while uncontracted forms vary pretty much equally between an immediately post-participle position and some other position (42 for facta est vs. 40 for est facta, facta...est, est...facta). A chi-square test reveals this distribution to be statistically significant, though the strength of the effect is somewhat weaker than that for masculine participles as indicated by the \( \phi \)-coefficient (\( \chi^2 = 24.4769, p = 7.52e-07, \phi = 0.1834596 \)). Thus, non-clitic *est* was clearly distributionally freer than its clitic counterpart.

Another interesting observation is that -st is rarely positioned in 2P in the clause (in 5% of instances), whereas the full variant *est* can occur in this position much more readily (in 29% of instances) (Pezzini, 2012: 181). This latter observation, as Pezzini (2012: 262) himself notes, contradicts what we usually find with clitics — namely that they seek out exactly those positions in the clause that -st avoids in 95% of instances. However, Pezzini does not explicitly define 2P; from the examples he gives it seems to mean second linear position in the clause. We will see below that this is problematic.

### 9.2.2.3 Interim summary

We have seen considerable evidence for clitic forms as opposed to non-clitic forms varying with respect to serialization. In otherwise non-configurational languages, clitics are constrained to appear in certain positions. In configurational languages, too, there is evidence that their syntax differs from that of their non-clitic cousins. Finally, we have observed an influence of the distinction of not only the clitichood of the auxiliary but also its prosodic nature on serialization of AVCs in a number of languages, including claims for Latin. In summary, clisis has a very prominent bearing on word order.

---

\( ^9 \) See below for a discussion of prosodic tree structure.

\( ^10 \) We have to be very careful about using terms such as *stress* for Latin, as it has been debated what type of accent it had in Classical times (see e.g. Allen 1973). I thus use the more neutral term *prosodic prominence* which can be applied generally to stress, pitch, and tone type systems.
Against this backdrop, we turn in the next section to a discussion of the variables whose effect we shall probe in this investigation and their operationalization.

9.2.3 Variable profile and operationalization

9.2.3.1 Auxiliary’s clitichood

We have seen above evidence that at least two forms of the verb *esse* ‘to be’ in Latin — i.e. the second person singular present *es* and the third person singular present *est* — can have clitic allomorphs, with the variants -s and -st, respectively. I therefore configure a predictor for the auxiliary’s clitichood (AuxiliaryClitichood). A word of caution is required at this point. Although we cannot diagnose the *actual* clitichood by looking at a text, because, for instance, clisis is not represented orthographically in our textual editions (see the discussion above), we can instead code for the *potential* clitichood. Thus, if the auxiliary was an *es* or *est* word form, I tagged it as being clitic (which I emphasize is to be understood as ‘having clitic potential, given the evidence’); all other word forms (e.g. *sunt, erant,...*) are tagged as being nonclitic (again to be understood as ‘probably not having clitic potential, given the evidence’).

This annotation was based on the auxiliary’s orthographic form, and so was performed automatically via R scripts.

```r
pac$AuxiliaryClitichood <- pac$Auxiliary
levels(pac$AuxiliaryClitichood)[levels(pac$AuxiliaryClitichood) == "es"] <- "clitic"
levels(pac$AuxiliaryClitichood)[levels(pac$AuxiliaryClitichood) == "est"] <- "clitic"
levels(pac$AuxiliaryClitichood)[levels(pac$AuxiliaryClitichood) != "clitic"] <- "nonclitic"

levels(pac$AuxiliaryClitichood)
[1] "nonclitic" "clitic"
```

All in all, there are 1503 auxiliaries with the level nonclitic and 906 auxiliaries with the level clitic spanning the n = 2409 observations.

9.2.3.2 Position of the PAC

We saw above evidence that clitics disfavour initial positions (termed 1P), seeking out second positions (termed 2P) in the clause after a prosodically strong host. If this is the case with respect to the PAC and if *esse* has clitic properties (at least, in some of its forms), we might expect that if the first element of the PAC is in 2P, it will be the auxiliary, thus resulting in an *est factus* serialization; by contrast, if the first element of the PAC is in 1P, it will be the participle, resulting in a *factus est* serialization. The reasoning behind this is obvious: putting the PAC in 1P when the first element is a clitic auxiliary will result in an ill-formed structure, while putting it in 2P when the first element is a clitic auxiliary will place the auxiliary in a position that is comfortably occupied by clitics cross-linguistically. In order to explore this possibility, I will establish a variable that codes for the position of the first element of the PAC (PositionOfPAC). This predictor is not straightforward to operationalize. Before we can proceed, we run up against several problems.

9.2.3.2.1 Motivating a prosodic definition of 2P

The first problem we encounter at the outset is defining what 2P exactly is.\footnote{Or, put round the other way, we could alternatively ask what counts as 1P.} So far we have only been talking about it loosely. For instance, is it to be defined with respect of the clause, the sentence, or some other formal structural unit? In what follows we will briefly discuss various approaches that have sought to establish an answer to this question.

Early accounts of clisis, e.g. Wackernagel (1892), defined the position of clitics in linear terms: the clitic occupies second position in the clause, i.e. after the first full word. Wackernagel (1892) cites many examples of this from Ancient Greek, Sanskrit, and Latin. We have seen such examples above, and
clitics occurring in this position have been reported for non-Indo-European languages, such as Tagalog (Austronesian) (Spencer & Luís, 2012). However, in the following examples, the above Wackernagelian definition of 2P does not seem to hold at all, as in the following examples from Warlpiri.

(10) a. Wat-ngki manu karnta-ngku ka lu mgarri-rni
   man-ERG and woman-ERG TENSE 3PL.SUBJ tell-NPAST
   ‘Men and/or women tell (him off)’

b. Ngulya-ngka jinta-ngka ka lu paka-rni
   burrow-LOC one-LOC TENSE 3PL.SUBJ hit-NPAST
   ‘They kill (them) in the one burrow’

(10) (examples from Spencer & Luís 2012: 55, ex. 37b, c; originally from Laughren 2002: 95)

In these examples the clitic cluster ka lu does not follow the first word, but the first constituent. So, rather than being the first orthographic word, 2P might instead be defined as occurring after the first syntactic constituent (XP). As Spencer & Luís (2012: 41) notes, “[s]ubsequent work on clitics has shown that, if anything, it is more common for second-position clitics to be limited to the position after the first phrase in the clitic’s domain...”. These positions are termed by Halpern (1995: 15) as 2W (“second-word”) and 2D (“second-daughter”), respectively.

A further problem with the notion of 2P, even on the above definition, is that some entities do not count as initial. Consequently, in a wide range of languages which otherwise exhibit 2P clitics, the clitic actually appears in a later clausal location, such as in third or fourth place. Thus, as Zwicky & Halpern (1996: 10) note, such initial constituents “don’t count’ or are ‘invisible’ in determining second position”.

We see this in the following Ancient Greek example, taken from Goldstein (2010: 2).

(11) áněu gär dě mágou où sphi vómou esti thusčas pojččesthais
   without for indeed magus.GEN NEG them.CL.innom.is.3S.PR.IND sacrifices.ACC make.INF
   ‘For without a magus it is not licit for them to perform sacrifices’ (Hdt. 1. 132. 15)

In the above example the prepositional phrase áněu gär dě mágou ‘without a magus’ does not count as initial. The pronominal clitic sphi is attached to the negative adverb où instead. Another example, this time from Serbo-Croatian (from Bošković 2001: 67), is given in (12).

(12) Njegovom najboljem prijatelju prodali su knjigu.
   his best friend.DAT sold are book.ACC
   ‘To his best friend they sold the book.’

Here, the clitic auxiliary su is positioned not after the first full constituent njegovom najboljem prijatelju ‘to his best friend’, but after the second, namely the lexical verb prodali. The phenomenon we witness in (11) and (12) is known as “skipping” or “delaying” in the literature (cf. e.g. Zwicky & Halpern 1996; Bošković 2001).

There is also variability — both between and within individual languages — with respect to certain elements: sometimes they appear to count as 1P and the clitic can “see” it and thereby latches on to it, whereas in other cases it does not count and the clitic attaches to a host later in the clause (cf. Zwicky & Halpern 1996: x–xi). A case in point is that of the complementizer category, c. While clitics follow the complementizer in Serbian/Croatian, in Chamorro they occur in third position, that is after the phrase following the complementizer (Spencer & Luís 2012: 237; on the latter they cite Chung 200312).

In Latin complementizer environments, weak pronouns that are allegedly clitic13 can appear after the...
complementizer or after the first constituent (Devine & Stephens, 2006), as the following alternation shows.

(13) a. cum tibi senatus ex aerario pecuniam
when you.DAT senate.M.SG.NOM from treasury.N.SG.ABL money.F.SG.ACC
prompsisset
withdraw.3.SG.PLPRF.ACT.SBJ
‘when the senate had withdrawn the money from the treasury for you’ (Cic. Verr. 2.3.195)
b. cum senatus tibi quaternos HS
dederit
give.3.SG.PRF.ACT.SBJ
‘when the senate has given you a price of four sesterces’ (Cic. Verr. 2.3.196)
(Examples from Devine & Stephens 2006: 294)

In (13a), the pronominal clitic tibi attaches to the complementizer cum, but in (13b) it is delayed and placed after the first full NP constituent senatus. There is thus variability with respect to clitic placement after complementizer type elements in Latin, and indeed in other languages.

The point to be taken home from the above discussion is that 2P is difficult to define, at least in linear terms. We will therefore define it in terms of the prosodic phonology of the clause, an approach that has been taken for a number of languages, including a.o. Serbian/Croatian (Radanović-Kocić, 1988, 1996; Bošković, 2001), Ancient Greek (Devine & Stephens, 1994; Goldstein, 2010), Sanskrit (Hock, 1996), and to some extent Latin (Devine & Stephens, 2006: 29–30 for general discussion; with reference to the auxiliary, see 191–193; with reference to clitic pronouns, see 277–312).

The theory of prosodic phonology (e.g. Selkirk 1984; Nespor & Vogel 1986) recognizes several layers of phonological representation, termed the prosodic hierarchy. An abridged version of this hierarchy is presented in (14).

(14) Intonational Phrase (Φ)
    | Phonological Phrase (φ)
    | Prosodic Word (ω)

A prosodic word typically corresponds to a single content (≡major) word, which we might class as nouns, verbs, adjectives, and adverbs (cf. Selkirk, 1984: 421, n.36). Moving up the hierarchy, a phonological phrase is defined as “any level of prosodic constituent structure that may include one or more major category words” (Selkirk, 1984: 29). It follows that an intonational phrase consists of one or more phonological phrases. Characteristic of the intonational phrase is its “comma-like” intonation, setting the phrase off from the rest of the sentence.

This prosodic chunking is not simply a theoretical construct. There is empirical evidence for it in that...
pauses tend to be longer higher up on the hierarchy, with relatively longer pauses at the left and right edges of intonational phrases than for instance at the left and right edges of phonological phrases.

To present an example of this hierarchical structure, the sentence *Dan baked the cake* can be prosodically parsed as follows.

(15) \(((\text{Dan} \omega) \phi (\text{baked} \omega (\text{the cake}) \omega) \phi) \Phi\)

While prosodic constituents are not isomorphic with syntactic constituents (see a.o. Selkirk 1984; for Dutch, see Booij, 1996: 219–220), there is a rough parallel: prosodic words are largely x\(^{th}\)s, phonological phrases are roughly equivalent to xps, and intonational phrases exhibit clause-like characteristics. That they are not isomorphic can be exemplified with English auxiliary ‘s-cliticization. As the following example demonstrates, although ‘s died forms a syntactic constituent (possibly inflectional phrase, ip, in mainstream theories of generative syntax), it is in the same prosodic contour as the subject, namely a phonological phrase. To my knowledge, no theory of grammar has suggested that a subject and an auxiliary form a syntactic constituent.

(16) a. [[Dan]NP [‘s died]IP]S ← syntactic parse  
   b. (((Dan’s) \omega) (died) \omega) \Phi ← prosodic parse

Such a theory can be, and has been, used to define 2P explicitly: the clitic occupies the second position in the intonational phrase that contains it.\(^{19}\) While the clitic can attach onto prosodic words and intrude into phonological phrases in its local domain, it cannot jump out of that local intonational domain and enter one further above it. An attempt has been made to formalize this scenario in (17).

(17) a.  
   \[
   \Phi_1 \downarrow \Phi_2 \downarrow \Phi_1 \downarrow \Phi_2 \downarrow \omega_1 \omega_2 \omega_3 \omega_4 \text{clitic}
   \]

b.  
   \[
   \Phi_1 \downarrow \Phi_2 \downarrow \omega_1 \omega_2 \omega_3 \omega_4 \text{clitic}
   \]

c.  
   \[
   \Phi_1 \downarrow \Phi_2 \downarrow \omega_1 \omega_2 \omega_3 \omega_4 \text{clitic}_i \quad \Phi_1 \downarrow \Phi_2 \downarrow \omega_1 \omega_2 \omega_3 \omega_4 \text{clitic}_i
   \]

d.  
   \[
   \Phi_1 \downarrow \Phi_2 \downarrow \omega_1 \omega_2 \omega_3 \omega_4 \text{clitic}_i \quad \Phi_1 \downarrow \Phi_2 \downarrow \omega_1 \omega_2 \omega_3 \omega_4 \text{clitic}_i
   \]

The consequence of this for examples such as (11) and (12) above, where the clitic is not in second position, is that, because there is an intonational phrase boundary preceding the entity in phrasal-second position (or word-fifth and word-fourth, respectively), there is no phonological means by which the clitic can pass through this boundary. Suppressing the lower prosodic level bracketing, we obtain the following parses:

\(^{18}\)Again, this is more difficult to delineate for Latin, given that the existence of xPs for this language is not strong.
\(^{19}\)This is the regular formulation. We will see later that there may be evidence for clitics associated with different prosodic domains.
(18) (´ aneu g` ar d` e m´ agou)Φ (ou sphì vômos estì thusìas poiéssthai)Φ.

(19) (Njegovom najboljem prijatelju)Φ (prodali su knjigu)Φ.

With respect to elements such as conjunctions and prepositions that typically don’t host the clitic, this is because they are themselves prosodically weak and cannot host the enclitic. Instead, these proclitic entities are prosodically phrased with what follows. With respect to the variable complementizers we saw in (13a) and (13b) one would assume that in the former case the complementizer has prosodic definition, while in the latter case it does not. Consequently, in the former it counts because it is “stressed”, while in the latter condition it is proclitic and cannot host the pronominal clitic. This excursus into the theory of phonological phrasing and now allows us to define what we mean by 2P for operational purposes.

(20) Definition of 2P: 2P is the second position (whether after the first prosodic word or the first phonological phrase) in the clitic’s local intonational phrase.

9.2.3.2.2 Operationalization of prosodic 2P for a dead language This brings us to a second operational problem: how do we define prosodic phonological phrasing for Latin when we have no acoustic material to sample? In order to do so, I will simply pool the literature on intonational phonological phrasing and clisis cross-linguistically, and simply stipulate the environments which I consider to be a 2P position with respect to this domain. It is not foolproof, but, I submit, it is replicable across studies, and allows us to make some speculative, but suggestive, initial observations. I will also give some examples from the dataset, but without at this point reflecting on the word order.

• Preceding sentence. By definition, a preceding sentence constitutes a separate intonational phrase and utterance phrase (U) in itself.20 An example is given in (21).

(21) a. at Germani celeriter ex consuetudine sua
but Germans.m.nom.pl swiftly from custom.f.abl.sg own.f.abl.sg
phalange facta impetus gladiorum
phalanx.f.abl.sg make.ptcp.pl prf.f.abl.sg attack.m.acc.pl swords.m.gen.pl
exceperunt. reperti sunt complures nostri
miličes our.m.pl.nom soldiers.m.pl.nom
‘But the Germans swiftly formed a phalanx, as was their custom, and took the sword-attack. Very many of our own soldiers were found to ’ (Caes. Gal. 1. 52. 4)

b. U(Φ(at Germani celeriter ex consuetudine sua phalange facta impetus gladiorum exceperunt.)Φ)U(Φ(reperti sunt complures nostri miličes . . . )Φ)U

• Participle phrases. Radanović-Kocić (1996: 440), discussing Serbian/Croatian, observes that “participial constructions obligatorily constitute a separate [intonational phrase], i.e. they are pronounced with a break...which separates them intonationally from the main clause”. Importantly, participle phrases have also been argued to constitute separate intonational phrases in Latin by Devine & Stephens (2006: 292), who note that “[t]his applies both when they are clausal circumstantial in the ablative absolute...and when they are attached to participant noun phrases”.21
Two examples, one with the participle phrase in ablative absolute function and a second with the participle phrase modifying a clausal argument (here a null subject), are given in (22) and (23), respectively.

(22) a. bello confecto cognitum
    war-N.SG.ABL settled-PTCP.PL.PRF.N.SG.ABL discovered-PTCP.PL.PRF.N.SG.NOM
    est
    be-3SG.PRS.IND
    ‘After war had been settled, it was discovered that…’ (Caes. Civ. 3. 60. 4)

b. Φ(bello confecto)Φ Φ(cognitum est)Φ

(23) a. in proelio pugnans aduersus Vettones
    in battle-N.SG.ABL fighting-PTCP.PRS.M.SG.NOM against Vettones-M.PL.NOM
    occisus est
    killed-PTCP.PRF.M.SG.NOM be-3SG.PRS.IND
    ‘He was killed in battle fighting against the Vettones’ (Nep. Ham. 4. 2)

b. Φ(in proelio)Φ Φ(pugnans aduersus Vettones)Φ Φ(occisus est)Φ

- Relative clauses. According to a number of researchers relative clauses also typically phrase into a separate intonational domain (cf. e.g. Nespor & Vogel, 1986: 188; Zwicky & Halpern, 1996: xiii; Hock, 1996: 207; Radanović-Kocić, 1996: 437; Dehé, 2009: 570). A Latin example is given in (24).

(24) a. sic bellum, quod rex aduersus Datamen susceperat, sedatum
    in.this.way war-N.SG.NOM, REL.N.SG.ACC king-M.SG.NOM against Datames-M.SG.ACC take.up-3SG.PL.PRF.IND, calm-PTCP.PRF.N.SG.NOM
    est
    be-3SG.PRS.IND
    ‘In this way, the war, which the king had taken up against Datames, was calmed’ (Nep. Dat. 8. 6)

b. Φ(sic)Φ Φ(bellum)Φ Φ(quod rex aduersus Datamen susceperat)Φ Φ(sedatum est)Φ

- Adjunct/argument clauses. According to Devine & Stephens (2006: 293) adjunct and argument cps phrase into a separate intonational domain, and consequently cannot host (pronominal) clitics. (I assume this refers to both finite and non-finite cps.)

(25) a. reliquos=que ut idem facerent
    rest-M.PL.ACC=and that same-N.SG.ACC do-3PL.IMPF.SBJV
    hortatus est
    encourage-PTCP.PRF.M.SG.NOM be-3SG.PRS.IND
    ‘And he encouraged the rest to do the same’

b. Φ(reliquos=que)Φ Φ(ut idem facerent)Φ Φ(hortatus est)Φ

---

22 An alternative prosodic parse for this sentence would be as follows: Φ(in proelio)Φ(pugnans aduersus Vettones)Φ occisus est)Φ.

23 It has been claimed that this only concerns non-restrictive (roughly similar to supplementary) relative clauses, not restrictive (roughly similar to integrated) ones; see e.g. Radanović-Kocić (1996: 437), who shows that Serbian/Croatian clitics are sensitive to the distinction. However, it is sometimes very difficult to tell between the two types. The example in (24) could be interpreted restrictively, where the relative clause defines the war, or non-restrictively, where the identity of the war is not at issue and the relative clause merely adds supplementary information.
• Parentheticals. These are another category that get phrased into their own intonational domains (Selkirk, 1984: 295; Nespor & Vogel, 1986: 188; Hock, 1996; Radanović-Kocić, 1996; Bošković, 2001; Devine & Stephens, 2006; Dehé, 2009: 570).

(26) a. cui quidem hodierno die, patres
   REL.M.SG.DAT indeed of.today-M.SG.ABL day-M.SG.ABL, father-M.PL.VOC
   conscripti — nunc enim primum ita conuenimus ut
   conscript-M.PL.VOC — now for first so meet-1PL.PRF.IND that
   illius beneficio possemus ea
   DEM.M.MGEN benefit-N.SG.DAT be.able-1PL.IMPF.SBJV DEM.N.PL.ACC
   quae sentiremus libere dicere — tribuenda
   REL.N.PL.ACC feel-1PL.IMPF.SBJV freely say-PRS.IND — grant-GER.F.SG.NOM
   est
   be-3SG.PRS.IND authority-F.SG.NOM
   ‘Today, Conscript Fathers — for now for the first time we have gathered in such a way that, thanks to him, we are able to say what we feel freely — authority is to be granted to him’ (Cic. Phil. 3. 5)

b. Φ(cui quidem hodierno die)Φ Φ(patres conscripti)Φ (— nunc enim primum ita conuenimus ut illius beneficio possemus ea quae sentiremus libere dicere —)Φ Φ(tribuenda est auctoritas)Φ

• Supplementary appositives are yet another category (cf. Hock, 1996; Radanović-Kocić, 1996; Bošković, 2001: 66; Devine & Stephens, 2006; Dehé, 2009: 570).

(27) a. cum Athenis, splendidissima ciuitate,
   since Athens-F.PL.ABL, splendid-SUPL.F.SG.ABL state-F.SG.ABL,
   natus esset
   be-born-PTCPL.PRF.N.SG.NOM be-3SG.IMPF.SBJV
   ‘Since he had been born in Athens, a most splendid state’ (Nep. Alc. 11. 2)

b. Φ(cum Athenis)Φ Φ(splendidissima ciuitate)Φ Φ(natus esset)Φ


(28) a. scelerate, inuentus qui . . .
   wicked-M.SG.VOC, find-PTCPL.PRF.M.SG.NOM be-2SG.PRS.IND REL.M.SG.NOM . . .
   ‘You criminal, you were found to . . . ’ (Cic. Phil. 2. 85)

b. Φ(scelerate)Φ Φ(inuentus qui . . . )Φ

In addition to the above, several other environments can delay the appearance of 2P clitics.

• Fronted constituents. It has been argued that fronted constituents are phrased into their own intonational domain (Nespor & Vogel, 1986: 188; Aissen, 1992; Hock, 1996; Bošković, 2001: 66). Along these lines, Goldstein (2010: 12, cf. 81), writing on Ancient Greek clitics pronouns, observes that “clause-initial ‘preposed’ noun phrases might knock a clitic out of clause-second position: but . . . it is still second within an intonational unit”. Deciding what and what is not “fronted” in Latin

24It has been argued that clause-final vocative can be prosodically restructured into the intonational phrase of the nuclear clause (see Hock, 1996: 208). It should be noted that, however, this is irrelevant to our analysis, as we are interested in the preceding sentential context.
9.2. CLISIS AND PROSODIC STRUCTURE

is not straightforward, as it allows largely unconstrained word orders to occur. One might therefore ask — what is fronted with respect to what? I will bypass this question here, and simply say that if constituents $\alpha$ and $\beta$ are displaced with respect to each other in accordance with one word order template proposed for Latin, i.e. that in (29), then fronting will be said to have occurred.

(29)  **Subject > Direct Object > Indirect Object/Oblique > Adjunct > Goal/Source > Non-Referential Direct Object**  
(Devine & Stephens, 2006: 79)

To give an example, in (30) the subject NP argument *tanta rerum commutatio* is pushed into a linear second place by the adjunct NP *horum adventu*; similarly in (31) the adjunct NP *eo die* is fronted with respect to the object NP argument *pace*. I assume that the non-fronted arguments are intonationally phrased with the *pac*.

(30)  a. horum adventu tanta rerum est facta

b. $\Phi$(horum adventu)$\Phi$ $\Phi$(tanta rerum commutatio est facta)$\Phi$

(31)  a. eo die pace sunt

b. $\Phi$(eo die)$\Phi$ $\Phi$(pace sunt usi)$\Phi$

- *Heavy phrases.* It has been suggested that heavy or branching NPs can form an intonational phrase, and thus delay clitic placement (Devine & Stephens, 2006: 192), as in the following Serbian/Croatian example.


b. $\Phi$(Jezičke razine više od rečenice)$\Phi$ vrlo =su . . .

An example from Latin is given by Devine & Stephens (2006: 192), which I repeat below in a modified form as (33).

(33)  a. signa=que militaria ex proedio ad standard-n.pl.nom=and military-n.pl.nom from battle-n.sg.abl to Caesarem sunt relata clxxx

b. $\Phi$(signaque militaria ex proedio)$\Phi$ $\Phi$(ad Caesarem sunt relata clxxx)$\Phi$

25I have to confess that in example (33), it is not clear to me whether the prepositional phrase *ex proedio* ‘from the battle’ is a modifier in the structure of the subject NP, as Devine & Stephens (2006) syntactically bracket it, or whether it is an adjunct to the verb phrase.
Drawing on the above, the following positional categories were established:

- The level 1p is reserved for PACs that occur in absolute initial sentence position or immediately follow any of the following types of categories: participle phrases, relative clauses, adjunct and argument clauses, parentheticals, supplementary appositives, and vocatives \( (n = 327) \).

- The level 2p is reserved for PACs that occur after the first constituent which is in absolute initial sentence position or the constituent follows any of the following types of categories: participle phrases, relative clauses, adjunct and argument clauses, parentheticals, supplementary appositives, and vocatives \( (n = 307) \).

- The level typical concerns other positions that clitics typically occupy: immediately after a higher verb boundary, following two switched constituents (one being fronted with respect to the other, see above), and following two constituents \( \alpha \) in \( \beta \) in which \( \alpha \) contains at least one more word than \( \beta \) \( (n = 401) \).

- The level non-typical concerns other positions that clitics do not typically occupy: following two constituents \( \alpha \) and \( \beta \) in which \( \alpha \) contains at least one less word than \( \beta \), immediately after a discourse connective or coordinator, and in a position that is non-adjacent to the higher verb \( (n = 362) \).

- The level other is the default category, if the observation does not meet any of the above criteria \( (n = 1012) \).

### 9.2.4 Statistical modelling

#### 9.2.4.1 Auxiliary cliticood

Evaluating first the predictor of auxiliary cliticood, there is, overall, a striking bias between potential clitic and non-clitic forms. Forms of esse ‘to be’ that could be clitics preferentially associate with factus est, whereas there is no apparent preference for non-clitic forms. This information is presented in Table 9.1 and graphically in the barplot in Figure 9.1.

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nonclitic</td>
<td>754 (50.17)</td>
<td>749 (49.83)</td>
</tr>
<tr>
<td>clitic</td>
<td>205 (22.63)</td>
<td>701 (77.37)</td>
</tr>
</tbody>
</table>

Table 9.1: PAC variant cross-classified by the cliticood of the auxiliary

While a straightforward \( \chi^2 \)-test of independence reveals this distribution to be statistically significant \( (\chi^2 = 177.7674, df = 1, p < 2.2e - 16) \), it is necessary to apply relevant statistical models. Thus, several models are constructed and compared. First, a varying intercepts model is constructed with the predictor for auxiliary cliticood:

```r
clitic.ri<-glmer(RealizationOfPAC~(1|LemmaParticiple)+(1|Text)+AuxiliaryCliticood,pac,family=binomial)
```

Varying slopes are then introduced. The second model specifies the fixed predictor of auxiliary cliticood, varying intercepts for text and verb lemma, and varying slopes for text and verb lemma:

```r
clitic.rs.both<-glmer(RealizationOfPAC~(AuxiliaryCliticood|LemmaParticiple)+(AuxiliaryCliticood|Text)+AuxiliaryCliticood,pac,family=binomial)
```

In the third model, a term is configured for the cliticood predictor, varying intercepts for text and verb lemma, and varying slopes for just the verb lemma:
9.2. CLISIS AND PROSODIC STRUCTURE

Figure 9.1: Distribution of PAC variants according to the auxiliary’s clitichood

clitic.rs.lemma<-glmer(RealizationOfPAC~(AuxiliaryClitichood|LemmaParticiple)+
(1|Text)+AuxiliaryClitichood,pac,family=binomial)

In the fourth and final model, a term is configured for the auxiliary clitichood predictor, varying intercepts for text and verb lemma, and varying slopes for just the text:

clitic.rs.text<-glmer(RealizationOfPAC~(1|LemmaParticiple)+
(AuxiliaryClitichood|Text)+AuxiliaryClitichood,pac,family=binomial)

The four models are then compared with respect to their deviance and AIC values. The summary output is reported in Table 9.2. As we can see the three more complex models are more informative than the model with just the varying intercepts (clitic.ri), as the differences in deviances from 2972.10 are significant on the relevant degrees of freedom. Finally the AIC value of the model named clitic.rs.both indicates that it might be the most informative and parsimonious. To check this we can compare its deviance (−2LL) with that of the two nested models, clitic.rs.lemma and clitic.rs.text. Given that the deviance difference between clitic.rs.both and clitic.rs.lemma is statistically significant (χ² = 35.259, df = 2, p = 2.206e-08) and that the deviance difference between clitic.rs.both and clitic.rs.text (χ² = 10.596, df = 2, p = 0.005002), the AIC value of the most complex model is confirmed. This suggests that there are robust and reliable interactions between auxiliary clitichood and both verb lemma and text. However, note that because the former model results in a greater deviance difference than the latter model, we can be slightly more confident about the robustness of clitichood × text interactions than clitichood × lemma interactions. That is, a larger deviance is obtained by omitting by-text slopes than by-lemma slopes.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>clitic.ri</td>
<td>4</td>
<td>2980.10</td>
<td>3003.25</td>
<td>-1486.05</td>
<td>2972.10</td>
</tr>
<tr>
<td>clitic.rs.lemma</td>
<td>6</td>
<td>2972.31</td>
<td>3007.04</td>
<td>-1480.16</td>
<td>2960.31</td>
</tr>
<tr>
<td>clitic.rs.text</td>
<td>6</td>
<td>2947.65</td>
<td>2982.37</td>
<td>-1467.83</td>
<td>2935.65</td>
</tr>
<tr>
<td>clitic.rs.both</td>
<td>8</td>
<td>2941.06</td>
<td>2987.35</td>
<td>-1462.53</td>
<td>2925.06</td>
</tr>
</tbody>
</table>

Table 9.2: ANOVA statistics for clitichood models

We should then compare the selected model (clitic.rs.both) with the baseline variance components model as established in Chapter 8. The ANOVA output, reported in Table 9.3, shows quite clearly that
the addition of the clisis predictor and the varying slopes significantly enhances the model. The \textit{AIC} and \(-2\text{LL}\) are both closer to zero, indicating that the resulting model is more parsimonious and explains more of the variance in the response variable than the baseline \textit{vcm}. We thus have statistical support, in addition to the theoretical support, of including a clisis predictor in the \textit{pac}'s information space.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>glmer.vcm.both</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clitic.rs.both</td>
<td>8</td>
<td>2941.06</td>
<td>2987.35</td>
<td>-1462.53</td>
<td>2925.06</td>
<td>208.35</td>
<td>5</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 9.3: ANOVA of baseline \textit{vcm} and \textit{glmer} containing auxiliary clitichood information

With the adequacy of the resulting model now established, let us walk through its parameters, given in Table 9.4. In terms of entering the binary variables into the model, for the predictor, non-clitic forms are dummy coded 0 and clitic forms are coded 1, while for the response variable, \textit{est factus} is dummy coded 0 and \textit{factus est} is dummy coded 1. Thus, the parameter estimates indicate the logit probability of a \textit{factus est} outcome given a one-unit change in the predictor, that is when it changes from a non-clitic form to a clitic-form. Specifically, the estimate indicates that when the type of auxiliary changes from a potentially non-clitic form (e.g. \textit{sunt, erat, . . .}), to one of a potentially clitic-form (e.g. \textit{es, est}), the logit probability of a \textit{factus est} outcome increases by 1.1861. This corresponds to an odds ratio of \(\exp(1.1861) = 3.274287\). The confidence intervals associated with this parameter range from 0.6604819 to 1.7116778 in logits, or in odds ratio terms from 1.9357249 to 5.538246. Given that the confidence intervals do not cross 0 in logits or 1 in odds, we can conclude that the population coefficient is accurately reflected by the present dataset. The significance of this parameter is also illustrated by dividing the estimate by its standard error to give a \(z\)-statistic of 4.423. Because this value is above the large sample critical \(z\)-value of 1.96, we have further confidence in the parameter estimate.

|                | Estimate | Std. Error | \(z\) value | Pr(>|\(z|)) |
|----------------|----------|------------|--------------|-------------|
| (Intercept)    | -0.1802  | 0.1690     | -1.066       | 0.286       |
| AuxiliaryCliticCoodClitic | 1.1861    | 0.2682     | 4.423        | 9.74e-06    |

Table 9.4: Final \textit{glmer} for serialization \(\sim\) auxiliary clitichood

It is also worth examining the by-text and by-lemma differences in the slopes for this particular predictor. The differences can be visualized in Figure 9.2 and Figure 9.3, respectively.

Evaluating the by-text slopes first, the slopes for \textit{B.Alex.} and Hirtius are very close to the overall grand mean slope. The overall slope underestimates for Caesar’s \textit{Gal.} and \textit{Civ.} and for Nepos; in these texts, clitichood has a more striking effect. The overall slope, by contrast, overestimates for four texts, namely \textit{B.Afr.}, \textit{B.Hisp.}, Cicero’s \textit{Phil.} and Varro; in these texts, clitichood does not have a particularly strong effect (all have positive slopes, however). Why should this be the case?

First, we have only coded for potential clitichood, taking the orthographic forms \textit{est} and \textit{es} as indicators thereof. This was because clinic auxiliaries are not represented in the editions we have been using, possibly as an ultimate result of copyists tending to ‘uncliticize’ the forms from -\textit{st} to the full variant \textit{est}, whilst retaining the original word order. It could be that in some texts (e.g. Nepos, Caesar \textit{Civ, Gal.}), clitic auxiliaries were very much present in the \textit{prima manus} (first hand manuscript) we do not have access to (hence clisis was more likely to have an effect), while in others (e.g. Cicero \textit{Phil.}), the clitic forms -\textit{st}, -\textit{s} may hardly have been represented in their \textit{prima manus} (hence clisis could not have the potential to have an effect). That ultimately could be down to author-based variability, reflecting the possibility that some authors were quite happy to use clisis in written texts, for instance in less-formal styles, while others shunned the use of clisis, for instance in more formal styles, much in the same way that some people use copula contraction in English written texts whereas other do not.

Related to this is the possibility that the more often a word form with clitic potential (e.g. \textit{est} or \textit{es}) is used in a text, the more likely it is to phonetically undergo reduction (we phonetically reduce
Figure 9.2: Slope adjustments for auxiliary clitichood by text

Figure 9.3: Slope adjustments for auxiliary clitichood by verb lemma
words that we use more often). Accordingly, some authors may use more forms with clitic potential (e.g. es, est) than other authors, hence the probability of a clitic form being used increases in the texts of the former writers. To explore this possibility, I plotted the individual by-text coefficient estimates as a function of the negative log probability of observing est/es vs. other forms of esse in those texts (see table 9.5 for the information that is plotted). Two outliers (Hirtius and Cicero’s Philippiks) were removed, as they were not conforming to the overall trend — as indicated in the left hand plot in figure 9.4.

With these removed, we have the scatterplot on the right, which indicates that as the \(-\log\) text probability of the clitic auxiliary goes up (that is, as the surprisal we see on getting an auxiliary increases, in Shannon’s information theoretic term), the estimate goes down, in support of this hypothesis. This pattern is confirmed by a regression analysis (\(\beta = -2.2209, SE = 0.4246, p = 0.003379\)). Again, this seems to suggest that the more likely a word form with clitic potential is, the more likely it is to be realized as an actual clitic (rather than in its full form), and consequently the more probable the factus est variant.

<table>
<thead>
<tr>
<th>Text</th>
<th>Estimate</th>
<th>(-\log) probability of es,est</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.Afr.</td>
<td>0.5746330</td>
<td>1.340624</td>
</tr>
<tr>
<td>B.Alex.</td>
<td>1.2551483</td>
<td>1.116961</td>
</tr>
<tr>
<td>B.Hisp.</td>
<td>0.5157716</td>
<td>1.476891</td>
</tr>
<tr>
<td>Caes. Civ.</td>
<td>1.8456580</td>
<td>1.094729</td>
</tr>
<tr>
<td>Caes. Gal.</td>
<td>1.6936206</td>
<td>1.016934</td>
</tr>
<tr>
<td>Cic. Phil.</td>
<td>0.5448940</td>
<td>0.875157</td>
</tr>
<tr>
<td>Hirt. Gal.</td>
<td>1.1788673</td>
<td>1.821612</td>
</tr>
<tr>
<td>Nep.</td>
<td>2.0216298</td>
<td>0.687617</td>
</tr>
<tr>
<td>Varr. RR</td>
<td>0.6299754</td>
<td>1.381899</td>
</tr>
</tbody>
</table>

Table 9.5: Individual by-text coefficient estimates and by-text clitic information content

Third, the difference may reflect differences in authors’ probabilistic grammars. All authors will have internalized a “rule” whereby clitic auxiliaries are placed last in the PAC. This is indicated by the fact that all the coefficients are positive in value, ranging from 0.5157716 in B.Hisp. to 2.0216298 in Cornelius Nepos. That is, there are no negative coefficients, which would mean that changing the auxiliary from a non-clitic form to a clitic form makes it more likely that we will observe est factus in the Late Republican Latin population. All authors associate cliticization of the auxiliary with factus est. However, and this is key, they do differ with respect to how they have internalized that rule, in that for some the constraint is more prominent and for others it is weaker. This difference, then, would be under the remit of idiolectal or sociolinguistic variability.

Let us now focus on lemma-based differences with respect to the slopes. Capio ‘capture’, cognosco ‘find out’, and cogito ‘think’ are all very close to the overall grand mean slope. However, the slope is underestimated for continco ‘contain’, demonstro ‘point out’, and adsentior ‘agree with’, in that clisis effects are stronger with these verb lemmas. By contrast, conor ‘try’, mitto ‘send’, and polliceor ‘promise’, for example, have overestimated slopes. Consequently, their individual slopes have to be adjusted downward. For those verb lemmas, clisis effects are weaker. In fact, while all texts displayed positive slopes, it should be noted that 2/604 verbs exhibit negative slopes: animadaerto ‘notice’ (\(-0.1193832\)) and conor ‘try’ (\(-0.4021799\)). This indicates that for these verbs clisis has a polar opposite effect to the majority of the verb lemmas in the dataset, that is clitic forms prefer associating with est factus and nonclitic forms with factus est. This can be seen in impressionistic terms by examining all the observations for conor and animadaerto (see Table 9.6). This shows that, for animadaerto, while only 2/7 instances of factus est have a potentially clitic auxiliary, 5/8 instances of est factus have a potentially clitic auxiliary, and

\[20\] We do this impressionistically here. However, for robustness, one would ideally need to use statistical diagnostics for outlier removal, such as Cook’s distance or Leverage values.
Figure 9.4: Relationship between by-text coefficient estimates (for serialization choice \( \sim \) auxiliary clitic- 
hood) and clitic information content. The upper plot contains outliers (Hirtius and Cicero’s *Phil.*), while 
in the lower plot the outliers have been removed and a regression line has been drawn.
for conor, while 3/14 instances of factus est have a potentially clitic auxiliary, 2/3 instances of est factus have a potentially clitic auxiliary. The trends observed here, while obviously not statistically significant in and of themselves \((p = 0.3147 \text{ for animaduerto and } p = 0.1912 \text{ for conor on Fisher exact tests})\), point to a differing behaviour for these lemmas compared with the rest. It is unclear why this should be so, however, and it would be unwise to make much of it (as we did for by-text differences) without collecting more datapoints for these and other lemmas.

<table>
<thead>
<tr>
<th>animaduerto ‘notice’</th>
<th>conor ‘try’</th>
</tr>
</thead>
<tbody>
<tr>
<td>animaduersa esset</td>
<td>conatus esset</td>
</tr>
<tr>
<td>animaduersum esset</td>
<td>conata esset</td>
</tr>
<tr>
<td>animaduersum esset</td>
<td>conatus est</td>
</tr>
<tr>
<td>esset animaduersum</td>
<td>conati sunt</td>
</tr>
<tr>
<td>animaduersum esset</td>
<td>conati sunt</td>
</tr>
<tr>
<td>esset animaduersum</td>
<td>conatum esse</td>
</tr>
<tr>
<td>animaduersum est</td>
<td>conati sunt</td>
</tr>
<tr>
<td>animaduersum est</td>
<td>conatus esset</td>
</tr>
<tr>
<td>esset animaduersum</td>
<td>es conatus</td>
</tr>
<tr>
<td>est animaduersum</td>
<td>conatus es</td>
</tr>
<tr>
<td>est animaduersum</td>
<td>conatus essem</td>
</tr>
<tr>
<td>est animaduersa</td>
<td>sunt conati</td>
</tr>
<tr>
<td>animaduersum esset</td>
<td>est conatus</td>
</tr>
<tr>
<td>est animaduentendus</td>
<td>conatus esset</td>
</tr>
<tr>
<td>sunt animaduersae</td>
<td>conati sunt</td>
</tr>
<tr>
<td></td>
<td>conatus est</td>
</tr>
<tr>
<td></td>
<td>conatus sit</td>
</tr>
</tbody>
</table>

Table 9.6: Observations on the verb lemmas animaduerto and conor

9.2.4.2 Position of the PAC

A first overview of the overall distribution of the PAC’s position in the prosodic phrase and its internal serialization is given in Table 9.7. The distribution is broadly in line with what we might expect, given the characterization of auxiliaries as clitics. When the PAC is in 1P in the prosodic phrase, the est factus serialization is strongly dispreferred, as that would put the auxiliary in an unsuitable position. The est factus serialization is also strongly dispreferred in non-typical clitic positions, but it somewhat more preferred, in proportional terms, in typical clitic positions and in 2P. This distribution is of course highly significant when submitted to a \(\chi^2\)-test of independence \((\chi^2 = 180.6118, df = 4, p < 2.2e - 16)\).

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1P</td>
<td>43 (13.15)</td>
<td>284 (86.85)</td>
</tr>
<tr>
<td>2P</td>
<td>162 (52.77)</td>
<td>145 (47.23)</td>
</tr>
<tr>
<td>typical</td>
<td>206 (51.37)</td>
<td>195 (48.63)</td>
</tr>
<tr>
<td>other</td>
<td>454 (44.86)</td>
<td>558 (55.14)</td>
</tr>
<tr>
<td>nontypical</td>
<td>94 (25.97)</td>
<td>268 (74.03)</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
</tr>
</tbody>
</table>

Table 9.7: PAC variant cross-classified by the PAC’s position in the prosodic phrase

We now move on to statistical modelling to evaluate the position of this predictor in the information space in an empirically more adequate way. Four models are constructed and compared. First, a varying intercepts model is fitted to the data:

```r
position.ri <- glmer(RealizationOfPAC~PositionOfPAC+(1|LemmaParticiple)+(1|Text), pac, family=binomial)
```
9.2. CLISIS AND PROSODIC STRUCTURE

Terms for varying slopes are then introduced sequentially. The second model fitted to the data has varying intercepts and varying slopes for both lemma and text.

\[
\text{position.rs.both<-glmer(RealizationOfPAC~PositionOfPAC+}
\]
\[
\text{(PositionOfPAC|LemmaParticiple)+(PositionOfPAC|Text),pac,family=binomial)}
\]

The third model has varying intercepts for both lemma and text, and varying slopes for only lemma.

\[
\text{position.rs.lemma<-glmer(RealizationOfPAC~PositionOfPAC+}
\]
\[
\text{(PositionOfPAC|LemmaParticiple)+(1|Text),pac,family=binomial)}
\]

The fourth and final model has varying intercepts for both lemma and text, and varying slopes for only text.

\[
\text{position.rs.text<-glmer(RealizationOfPAC~PositionOfPAC+}
\]
\[
\text{(1|LemmaParticiple)+(PositionOfPAC|Text),pac,family=binomial)}
\]

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>ri</td>
<td>7</td>
<td>2980.77</td>
<td>3021.28</td>
<td>-1483.39</td>
</tr>
<tr>
<td>lemma</td>
<td>21</td>
<td>2989.69</td>
<td>3111.21</td>
<td>-1473.84</td>
</tr>
<tr>
<td>text</td>
<td>21</td>
<td>2987.23</td>
<td>3108.75</td>
<td>-1472.61</td>
</tr>
<tr>
<td>both</td>
<td>35</td>
<td>2998.19</td>
<td>3200.73</td>
<td>-1464.09</td>
</tr>
</tbody>
</table>

Table 9.8: ANOVA statistics for position models

When we compare the least complex model (position.ri) with the three more complex models, none of the more complex models is significantly better (this information can be worked out from Table 9.8). That is, although their deviances are smaller than position.ri's deviance of 2966.77, the differences are not significant enough to be of explanatory utility. For example, the deviance difference between position.rs.both and position.ri is 38.588, which on 35−7 = 28 degrees of freedom is only marginally significant at \( p = 0.08778 \) on a \( \chi^2 \)-distribution (we normally only accept models at the \( p < 0.05 \) threshold). As such, model position.ri exhibits the lowest \( AIC \) (2980.77), and is the one to be inspected further. We then compare this model with the baseline vcm via an ANOVA, and find it to be of significantly better explanatory power (see Table 9.9).

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>glmer.vcm.both</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>position.r</td>
<td>7</td>
<td>2980.77</td>
<td>3021.28</td>
<td>-1483.39</td>
<td>2966.77</td>
<td>166.63</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 9.9: ANOVA of baseline vcm and glmer containing positional information

With the model now firmly established, let us now inspect the parameters of position.ri more closely. The model is reported in Table 9.10, along with confidence intervals in Table 9.11. When the position of the pac is in 1P compared with the default category other, the logit probability of a factus est outcome increases by 1.58905; this is equivalent to an odds of 4.899093 for every 1 time that it does not occur. When the position of the pac is in 2P compared with the default category other, the logit probability of a factus est outcome decreases by -0.37655; this is equivalent to an odds ratio of 0.6862248. When the position of the pac is in a typical clitic position, the logit probability of a factus est outcome, again, decreases by -0.35379; this is equivalent to an odds ratio of 0.7020224. Finally, when the position of the pac is in a non-typical clitic position, the logit probability of a factus est outcome increases by 0.74883; this is equivalent to an odds ratio of 2.114525. When all of the coefficients are divided by their standard errors, their absolute z-values are above the large sample critical value of 1.96, and we can thus conclude that they are statistically significant. The position of the pac in the prosodic phrase is of central importance in determining the internal serialization of the pac.
CHAPTER 9. PHONOLOGY

One question is immediately apparent. As noted above, Pezzini observed that clitic forms such as -st do not occur in 2P positions, or at least according to his unstated operationalization of it. It therefore might be that what we have been hitherto been calling potentially clitic forms (e.g. est) might differ from what we have hitherto been calling potentially nonclitic forms (e.g. erat) with respect to the influence of the pac’s position in the prosodic phrase. We now take this into account. The most effective way to see this is to construct two individual models, rather than introducing interaction effects at this stage. We will also just include varying intercepts rather than varying slopes, and not pursue the modelling procedure further than is necessary to get an overall indication of the trends present in the data. Table 9.12 presents a glmer for serialization choice modelled as a function of pac position when the auxiliary is of a potentially non-clitic form, and Table 9.13 presents a glmer for serialization choice modelled as a function of pac position when the auxiliary is of a potentially clitic form.

Table 9.12: glmer for serialization ~ position

| Estimate  | Std. Error | z value | Pr(>|z|) |
|-----------|------------|---------|----------|
| (Intercept) | 0.03595 | 0.15187 | 0.237 | 0.81289 |
| PositionOfPAC1p | 1.58905 | 0.18195 | 8.733 | <2e-16 |
| PositionOfPAC2p | -0.37655 | 0.13652 | -2.758 | 0.00581 |
| PositionOfPACtypical | -0.35379 | 0.12324 | -2.758 | 0.00581 |
| PositionOfPACnontypical | 0.74883 | 0.14016 | 5.343 | 9.16e-08 |

Table 9.11: Confidence intervals for glmer for serialization ~ position

<table>
<thead>
<tr>
<th>Logs</th>
<th>Odds</th>
<th>2.5%</th>
<th>97.5%</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.26</td>
<td>0.33</td>
<td>0.77</td>
<td>1.40</td>
<td></td>
</tr>
<tr>
<td>PositionOfPAC1p</td>
<td>1.23</td>
<td>1.95</td>
<td>3.43</td>
<td>7.00</td>
<td></td>
</tr>
<tr>
<td>PositionOfPAC2p</td>
<td>-0.64</td>
<td>-0.11</td>
<td>0.53</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>PositionOfPACtypical</td>
<td>-0.60</td>
<td>-0.11</td>
<td>0.55</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>PositionOfPACnontypical</td>
<td>0.47</td>
<td>1.02</td>
<td>1.61</td>
<td>2.78</td>
<td></td>
</tr>
</tbody>
</table>

Table 9.12: glmer for serialization ~ position | auxiliary = nonclitic

| Estimate  | Std. Error | z value | Pr(>|z|) |
|-----------|------------|---------|----------|
| (Intercept) | -0.2026 | 0.1709 | -1.185 | 0.23592 |
| PositionOfPAC1p | 1.3490 | 0.2228 | 6.055 | 1.41e-09 |
| PositionOfPAC2p | -1.0025 | 0.2147 | -4.670 | 3.02e-06 |
| PositionOfPACtypical | -0.60 | -0.11 | 0.55 | 0.89 |
| PositionOfPACnontypical | 0.6965 | 0.1653 | 4.213 | 2.52e-05 |

There are several striking features that emerge from the statistical models presented in Table 9.12 and Table 9.13. When the auxiliary is of a potentially nonclitic type, all of the coefficients are statistically significant. Specifically, for pac’s positioned in 1P in the intonational phrase or positioned in some other nontypical clitic locus that is unsuitable for clitics, then the logit probability of a factus est outcome is positive; however, when positioned in 2P in the intonational phrase or positioned in some other typical clitic position, then the logit probability of a factus est outcome decreases. In other words, when the pac is in 1P, the participle factus is placed before the auxiliary, so as to prevent the latter from being at the immediate left-edge of the intonational phrase; when the pac is in 2P, the auxiliary is in a suitable position, and the participle need not be brought into service to act as a prosodic buffer as another element/word/constituent already fulfills that function. What we have been labelling clitic auxiliaries appear to behave both similarly and differently at the same time. They are similar in that when the pac is in first position, the participle factus is placed before the auxiliary, so as to prevent the latter from being at the immediate left-edge of the intonational phrase. However, they are different in that pac’s in 2P nor those in other typical clitic positions do not behave significantly differently from those in the default position.
9.2. CLISIS AND PROSODIC STRUCTURE

|                      | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------------|----------|------------|---------|----------|
| (Intercept)          | 0.7984   | 0.2515     | 3.174   | 0.00150  |
| PositionOfPAC1p      | 1.7661   | 0.3651     | 4.838   | 1.31e-06 |
| PositionOfPAC2p      | -0.3341  | 0.2242     | -1.491  | 0.13609  |
| PositionOfPACtypical | -0.3023  | 0.2310     | -1.309  | 0.19059  |
| PositionOfPACnontypical | 0.9820  | 0.3232     | 3.038   | 0.00238  |

Table 9.13: glmer for serialization ~ position | auxiliary = clitic

In the next section we discuss the consequences of this distinction in positional behaviour.

9.2.5 Discussion

Let us now recapitulate the main findings concerning auxiliary cliticoid and its impact on PAC serialization. Auxiliaries that we have labelled as being “potentially clitic” favour *factus est* serializations, whereas those that are “potentially non-clitic” do not demonstrably appear to favour either serialization. Secondly, there is an effect of where the PAC is located in the local intonational phrase. In broad terms, when the PAC is in 1P in the intonational phrase, the *factus est* serialization is strongly preferred; when the PAC is in 2P in the intonational phrase, the *est factus* serialization is strongly preferred. However, this effect appears to be modulated depending on the type of apparent “cliticability” of the auxiliary, that is “potentially clitic” forms and “potentially nonclitic” forms differ with regard to how they behave in prosodic phrases, ultimately affecting PAC serialization. Specifically, “potentially clitic” forms are strongly repelled from 1P in the prosodic phrase, and consequently conspire to effect a *factus est* serialization, but there is no effect when the PAC is in 2P in the prosodic phrase. By contrast, “potentially nonclitic” forms are also strongly repelled from 1P, thereby effecting *factus est* serializations, but they are also attracted to 2P, thereby effecting *est factus* serializations when positional conditions are optimal. Should it not be the case that “potentially nonclitic” forms should be able to go anywhere, while “potentially clitic” forms should be barred from 1P and be attracted to 2P? What is going on?

In my view, this is suggestive of a complex auxiliary realization system in Latin. In basic terms, auxiliaries in Latin can be realized overtly (as they are for all datapoints in our sample) or they can be realized covertly, that is elided from the clause. When the auxiliary is overtly realized there are several options. The first option is for the auxiliary to receive an accent, e.g. [ˇest], much like Serbian/Croatian *jesu* (see above). If this is the case, then the auxiliary should be able to optionally follow the participle, and result in a *factus est* serialization, or precede it, and result in an *est factus* realization. The second option is for the auxiliary to undergo phonetic reduction, e.g. [ˇest] → [ˇst]. In this condition, it is an intonational phrase 2P clitic, just like pronominal *sphin* in Greek, the auxiliary *je* in Slovene, and the auxiliary *su* in Serbian/Croatian. The third option is for the auxiliary to be reduced to just the consonant cluster [-st]. Unlike [ˇst], [-st] is not an intonational phrase clitic, but rather one of a different sort — namely a *phonological* phrase clitic. That would mean, no matter where the PAC is positioned in the intonational phrase, it cannot get out of its own phonological phrase and thus has to remain in 2P within that domain. In this respect, [-st] would be essentially very similar to the Greek particle *ge*, which is a phonological phrase clitic. The three way contrast is, in my view, attractive in that it is both empirically and typologically supported. First, the different prosodic properties of the auxiliary (full ~ intonational phrase clitic ~ phonological phrase clitic) make sense of the serialization differences. Second, in typological terms, one only has to look at English, where the auxiliary *is* can be realized in three different ways, full as in *[iz]*, reduced as in *[az]*, or fully contracted as in *[z]* (MacKenzie, 2012).

What has been apparent in this section is that it has been crucial to delve into the properties of clitic auxiliaries in order to understand serialization differences. If no attempt is made to understand clisis, we cannot understand serialization differences in PACs. We will have no more to say about clisis effects until we move on to the multivariate statistical modelling in Part IV. However, while important, I would
content that clisis is not the only important predictor on the PAC alternation. In the next sections, we discuss some further phonological predictors.

9.3 Rhythmical Optimization

9.3.1 Introduction

This and the next main section discuss the effects of two phonological optimization principles — rhythmical and segmental optimization. Rhythmical optimization, the first principle we shall deal with, maintains that words should be formed of alternating strong and weak syllables; it is known as the Principle of Rhythmic Alternation. Specifically, a word with the rhythmical sequence [wsws] is optimal because it consists of a weak syllable, followed by a strong syllable, followed by a weak syllable, and then a final strong syllable: it is perfectly alternating. By contrast, a word with the rhythmical sequences of [swwws] and [swsss] are suboptimal, because in the former there is a lapse of strong syllables, while in the latter there are three strong syllables in adjacency.

9.3.2 Grammatical variability

There is some evidence that rhythmical optimization has an effect on the choice of grammatical alternatives. For instance, Jaeger (2006) examines variable complementizer/relative that. Given that that is typically unstressed Jaeger hypothesizes that “[i]f speakers can use optional that to avoid stress lapses . . . , optional that should therefore be dispreferred before and/or after unstressed syllables. . . Additionally, optional that maybe [sic] used to avoid a stress clash. . .”. Controlling for other factors, he finds some support of this hypothesis, with “only avoidance of lapses with following syllables [seeming] to affect optional that likelihood”.

Similarly, Schlüter (2003) investigates the distribution of the participial variants lit vs. lighted. Overall lit is more common (94%), but in single attributive uses it is the form lighted that dominates (86%). She observes that since most English nouns are accented on the first syllable, then if lit co-occurs with them, then the sequence will result in a stress clash (e.g. lit cándle), whereas if lighted co-occurs with them, then the optimal rhythmical structure is preserved where there exactly one unstressed syllable between the two accentual peaks (e.g. lighted cándle. The -ed suffix on the participle acts “as a stress clash buffer”.

By contrast, the form lit can be used to prevent a stress lapse when the noun is not accentuated on the first syllable: e.g. lit cigarées is preferred over lighted cigarées because in the latter there is a lapse of three unaccented syllables between the two accent peaks.27

Relevant for our purposes is that it can also have an effect on word order choices, as has been exemplified for the genitive alternation. Shih et al. (in press) demonstrate that “the Principle of Rhythmic Alternation. . . should be considered a potential influencer of construction choice in English. Its role is small, especially when compared to some semantic, pragmatic, processing, and other phonological factors, and rhythm, as a dependent on other predictors, does not have complete explanatory power of construction choices. But, though its role may be small, rhythmicity still participates in the decision between genitive construction alternatives.”

9.3.3 Rhythmical optimization in AVCs and PACs

There is to my knowledge no discussion of the effect of rhythmical/accentsual optimization on Latin PACs, but there is with respect to historical German AVCs. Sapp (2011), examining verb clusters in Middle High German and Early New High German, hypothesizes that if the preverbal expression is unstressed, an auxiliary-last sequence (2-1) will be more likely to follow; and if the preverbal expression is stressed, an

27Schlüter (2005) offers further case studies attesting to the reliability of rhythmic effects on the grammar.
9.3. RHYTHMICAL OPTIMIZATION

<table>
<thead>
<tr>
<th>Stressed nature of preceding word</th>
<th>auxiliary-last (2-1)</th>
<th>auxiliary-first (1-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>stressed</td>
<td>420 (54.47%)</td>
<td>216 (70.36%)</td>
</tr>
<tr>
<td>unstressed</td>
<td>351 (45.53%)</td>
<td>91 (29.64%)</td>
</tr>
<tr>
<td></td>
<td>771 (100%)</td>
<td>307 (100%)</td>
</tr>
</tbody>
</table>

Table 9.14: Sapp (2011: Table 4, 23, modified)

auxiliary-first (1-2) sequence will be more likely to follow. The reasoning behind this draws on earlier work (e.g. notably Behaghel 1932: IV, 87), in that specific sequences are preferred – e.g. heavy–light–heavy and light–heavy–light. For him, unstressed syntactic entities are taken to be pronouns, short adverbs, the negator, and da-compounds. For the Middle High German dataset, he finds that this hypothesis is borne out, as can be seen in table 9.14 ($p = 1.481\text{e-06}$, $OR = 0.504249, CI = [0.3752729, 0.6746009]$). In the Early New High German dataset, by contrast, there is no overall effect of preceding stress (Sapp, 2011: 60).

9.3.4 Operationalization and variable profile

Several ways of operationalizing rhythmical optimization were considered, including both multicategory and numerical predictors, which I discuss in this section.

First, I examined the primary accentual structure (i) across the boundary between the participle and the auxiliary in that sequence and (ii) across the boundary between the auxiliary and the participle in that sequence. To operationalize the placement of the primary word accent, I utilized the information in the standard handbooks (e.g. Kent 1932; Allen 1973, 1978) which themselves draw on the guides of the ancient grammarians (e.g. Quintilian). These guides note that in Classical Latin polysyllabic words a primary accent is placed on the penultimate syllable if it is heavy and on the antepenultimate syllable if the penultimate syllable is light: thus *profèctus* [wsw] but *pòsitus* [sww].28 In disyllabic words the primary accent falls on the first syllable whatever its quantity: e.g. *dátus* [sw]. The final syllable is extrametrical except in monosyllables where of course it is the only syllable. These rules are formulated in (34).29

\begin{equation}
\begin{align*}
\text{(34)} & \quad \text{a. } & L & H \ (X) \\
& & 1 & 2 \ (3) \\
& & \rightarrow & \text{accent 2: [wsw]} \\
\text{b. } & \quad L & L \ (X) \\
& & 1 & 2 \ (3) \\
& & \rightarrow & \text{accent 1: [sww]} \\
\text{c. } & \quad X \ (X) \\
& & 1 & (2) \\
& & \rightarrow & \text{accent 1: [sw]} \\
\text{d. } & \quad X \\
& & 1 \\
& & \rightarrow & \text{accent 1: [s]}
\end{align*}
\end{equation}

Using this information, I looked at the accentual structure across (i) the boundary between the participle and the auxiliary in that sequence and (ii) the boundary between the auxiliary and the participle in that sequence. I coded for whether the accentual structure resulted in (i) an accentual clash where two

28Here, ‘s’ = ‘accented’ and ‘w’ = ‘unaccented’. I ignore the issue of the type of accent (pitch or stress) the Classical Latin accent was, which is disputed in the literature, as this is irrelevant to the overall discussion.

29 ‘L’ = ‘light’; ‘H’ = ‘heavy’; ‘X’ = ‘light or heavy’.

115
accented syllables are juxtaposed (as in e.g. fact(um) est [ss] and est dátus [ssw]), (ii) a lapse where there is more than one unaccented syllable between the two accentual peaks (as in e.g. pósitus érat [swwsw] and éset accipiénda [swwsws]), or (iii) harmony where there is a single unaccented syllable between the two accentual peaks (as in e.g. dátus ést [sws] and éset rátæ) [ssw]). Here, then, there are two variables PartAuxRhythm1 and AuxPartRhythm1, both with three levels (clash vs. lapse vs. harmony). I term this group of variables Rhythm 1.

It is unclear whether probing just the primary accent is empirically sufficient. So I secondly looked at the primary and secondary accentual structure across (i) the boundary between the participle and the auxiliary in that sequence and (ii) the boundary between the auxiliary and the participle in that sequence. Primary accented syllables were operationalized using the above guide in (34). A word was considered to have a secondary accent on the initial syllable iff the syllable immediately following it was unaccented (Allen, 1973: 188–191): so, accipiénda has two accents (primary on the penultimate, secondary on the initial), but not *sèpósitus where the immediately post-initial syllable is accented. The variables were coded in exactly the same way as the preceding variables: again we have two variables PartAuxRhythm2 and AuxPartRhythm2, both with three levels (clash vs. lapse vs. harmony). This second group of variables is termed Rhythm 2.

The third and fourth operationalizations are numerical in nature rather than categorical and make use of Shih et al.’s (in press) measure of what they term eurhythmic distance (ED). This operationalization seeks to “[measure] how far away from perfectly alternating rhythm a given construction is” by taking into consideration the number of intervening syllables between the accent in the leftmost word in the construction and the accent in the rightmost word in the construction. It is calculated as in Equation (9.1):

\[ ED = |\# \text{ of unstressed syllables between the rightmost word accent and the leftmost word accent} - 1| \]

(9.1)

It follows that when the eurhythmic distance is “perfect” across the word border, then \( ED = 0 \). To exemplify, in actus ´ esset [ssws], there is exactly one unaccented syllable between the leftmost and the rightmost word accents; applying the \( ED \) measure here, we obtain \( |1 - 1| = 0 \). When the sequence is arrhythmic, we obtain \( ED \) values \( \geq 1 \) with the most arrhythmic values tending to infinity. As such, the measure penalizes lengthier accent lapses. To exemplify, in fácit(um) ´ est [ss] there is a clash of two accentual peaks; applying the \( ED \) measure here, we obtain \( |0 - 1| = 1 \). Similarly, in pósitus ´ ést [swws], there are two intervening unaccented syllables between the two accentual peaks; applying the \( ED \) measure here, we obtain \( |2 - 1| = 1 \). As a final example, in éset accipiénda [swwswsw] there are four intervening unaccented syllables between the two accentual peaks (assuming only the presence of a primary accent for the purpose of exemplification); applying the \( ED \) measure, we obtain \( |4 - 1| = 3 \).

Four variables were constructed using this measure. The first two variables, which I term EURHYTHMIC DISTANCE 1, concern the \( ED \) in (i) the factus est serialization (termed PartAuxED1) and (ii) the est factus serialization (termed AuxPartED1) using only primary accent information. The second two variables, which I term EURHYTHMIC DISTANCE 2, concern the \( ED \) in (i) the factus est serialization (termed PartAuxED2) and (ii) the est factus serialization (termed AuxPartED2) using both primary and secondary accent information.

While it is possible to discuss the models for all eight variables over the four variable groups, I decided to concentrate on the variables from the operationalization scheme that best characterizes rhythmical optimization. Specifically, I concentrate on the variables from Rhythm 2. There are several reasons for this. ED makes no distinction between lapses and clashes; this may be important, because it is known that the effects of lapses tend to be weaker than those of clashes. First, Rhythm 2 is to be preferred over

\[ ^{30} \text{However, it must be noted that the evidence of a secondary accent in Latin is less clear-cut than that for a primary accent (see below).} \]

\[ ^{31} \text{As one can see, the \( ED \) measure makes no distinction between clashes and lapses, as the authors themselves make explicit.} \]
9.3. RHYTHMICAL OPTIMIZATION

<table>
<thead>
<tr>
<th>Variable Group</th>
<th>Sum $AIC$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHYTHM 1</td>
<td>6434.6</td>
</tr>
<tr>
<td>RHYTHM 2</td>
<td>6394.8</td>
</tr>
<tr>
<td>EURHYTHMIC DISTANCE 1</td>
<td>6470.8</td>
</tr>
<tr>
<td>EURHYTHMIC DISTANCE 2</td>
<td>6483.9</td>
</tr>
</tbody>
</table>

Table 9.15: Model $AIC$s of various operationalizations of rhythmical optimization

RHYTHM 1 in that, while there is less evidence for secondary accents in the Classical Latin word, it is not entirely absent (see Allen 1973: 188–191). Second, I constructed individual GLMs for each of the eight variables. Model $AIC$s for both two variables within each variable group were summed together, which are reported in Table 9.15. As Table 9.15 shows, the variable group EURHYTHMIC DISTANCE 2 has the highest $AIC$ and the variable group RHYTHM 2 has the lowest $AIC$. Consequently, we prefer the variable group RHYTHM 2. Finally, the sheer volume of the variables involved means that a comprehensive report thereof will be prohibitively long. In sum, I thus feel justified in choosing the operationalization that is not only driven by the data, but it also backed up by the theoretical literature.

I hypothesize that *factus est* serializations will be preferred if (i) the accentual structure is harmonious across the boundary between the participle and the auxiliary in that sequence, (ii) the accentual structure produces a clash across the boundary between the auxiliary and the participle in that sequence, or (iii) the accentual structure produces a lapse across the boundary between the auxiliary and the participle in that sequence. I also hypothesize that *factus est* serializations will be dispreferred if (i) the accentual structure is harmonious across the boundary between the auxiliary and participle in that sequence or (ii) the accentual structure produces a clash across the boundary between the participle and the auxiliary in that sequence; in other words, in these final two conditions, *est factus* serializations will be the outcome of choice.

9.3.5 Statistical modelling

9.3.5.1 Participle/Auxiliary Rhythm

The overall distribution of this first rhythmic predictor is given in Table 9.16. As one can see, there does not appear to be much in the way of an asymmetry, with percentages fluctuating around 40% for *est factus* and 60% for *factus est*, and this seems to be supported by a $\chi^2$-test of independence ($\chi^2 = 2.10$, $df = 2$, $p = 0.35$).

<table>
<thead>
<tr>
<th>Serialization</th>
<th><em>est factus</em> (%)</th>
<th><em>factus est</em> (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>harmonious</td>
<td>499 (38.47)</td>
<td>798 (61.53)</td>
</tr>
<tr>
<td>clash</td>
<td>409 (41.40)</td>
<td>579 (58.60)</td>
</tr>
<tr>
<td>lapse</td>
<td>51 (41.13)</td>
<td>73 (58.87)</td>
</tr>
<tr>
<td>Total</td>
<td>959</td>
<td>1450</td>
</tr>
</tbody>
</table>

Table 9.16: Relationship between participle-auxiliary rhythm and serialization

Four models are now constructed and compared with ANOVAs. The first model contains the predictor and varying intercepts for lemma and text:

```
partauxrhythm.ri<-glmer(RealizationOfPAC~PartAuxRhythm2+(1|LemmaParticiple)+(1|Text),pac,family=binomial)
```

The second model contains the predictor along with both varying intercepts and varying slopes for lemma and text:

```
partauxrhythm.rs.both<-glmer(RealizationOfPAC~PartAuxRhythm2+(PartAuxRhythm2|LemmaParticiple)+(PartAuxRhythm2|Text),pac,family=binomial)
```
The third model contains the predictor, varying intercepts for lemma and text, and varying slopes for just lemma.

\[
\text{partauxrhythm.rs.lemma} \leftarrow \text{glmer(RealizationOfPAC~PartAuxRhythm2+ (PartAuxRhythm2|LemmaParticiple)+(1|Text),pac,family=binomial)}
\]

The third model contains the predictor, varying intercepts for lemma and text, and varying slopes for just text.

\[
\text{partauxrhythm.rs.text} \leftarrow \text{glmer(RealizationOfPAC~PartAuxRhythm2+ (1|LemmaParticiple)+(PartAuxRhythm2|Text),pac,family=binomial)}
\]

The relevant ANOVA output in Table 9.17. Both partauxrhythm.rs.both and partauxrhythm.rs.text supply significantly more information than the varying intercepts model partauxrhythm.ri \((\chi^2 = 21.95, \, df = 10, \, p = 0.01536 \text{ and } \chi^2 = 14.864, \, df = 5 \, p = 0.01096, \text{ respectively})\), whereas partauxrhythm.rs.lemma does not \((\chi^2 = 9.32, \, df = 5, \, p = 0.097)\). Furthermore, when the complex model partauxrhythm.rs.both is compared with the nested models in terms of its deviance reduction, it is significantly better than partauxrhythm.rs.lemma \((\chi^2 = 12.632, \, df = 5, \, p = 0.02709; \text{ recall, this includes only lemmatic information})\), and the added complexity is therefore justified, but not significantly better than partauxrhythm.rs.text \((\chi^2 = 7.086, \, df = 5, \, p = 0.2143; \text{ recall, this includes only textual information})\), and the added complexity is not justified. Consequently, it is the textual information that is doing all the work in reducing the deviance, and therefore partauxrhythm.rs.text must be adopted for further inspection. (Note further, that it has the lowest AIC in Table 9.17.)

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>partauxrhythm.ri</td>
<td>5</td>
<td>3141.44</td>
<td>3170.37</td>
<td>-1565.72</td>
<td>3131.44</td>
</tr>
<tr>
<td>partauxrhythm.rs.lemma</td>
<td>10</td>
<td>3142.12</td>
<td>3199.99</td>
<td>-1561.06</td>
<td>3122.12</td>
</tr>
<tr>
<td>partauxrhythm.rs.text</td>
<td>10</td>
<td>3136.57</td>
<td>3194.44</td>
<td>-1558.29</td>
<td>3116.57</td>
</tr>
<tr>
<td>partauxrhythm.rs.both</td>
<td>15</td>
<td>3139.49</td>
<td>3226.29</td>
<td>-1554.74</td>
<td>3109.49</td>
</tr>
</tbody>
</table>

Table 9.17: ANOVA statistics for various GLMERS containing rhythmic information on the participle/auxiliary boundary

This model is then compared with the baseline vcm in Table 9.18. This clearly shows that the addition of the predictor is important in explaining the variance in the response variable. The reduction in deviance of 16.83 \((3133.41 \, – \, 3116.57)\) is highly significant on \(7 \,(\,= \, 10 \, – \, 3)\) degrees of freedom.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.vcm</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>partauxrhythm.rs.text</td>
<td>10</td>
<td>3136.57</td>
<td>3194.44</td>
<td>-1558.29</td>
<td>3116.57</td>
<td>16.83</td>
<td>7</td>
<td>0.0185</td>
</tr>
</tbody>
</table>

Table 9.18: ANOVA of baseline VCM and selected GLMER containing rhythmic information on the participle/auxiliary boundary

With this information in hand, we can now inspect the parameters. The response variable is dummy (0-1)-coded, with factus est outcomes treated as successes (=1) and est factus outcomes treated as failures (=0). The predictor variable is also dummy coded, with the level harmonious established as the baseline. The parameter estimates are reported in Table 9.19.

None of the parameter estimates are important statistically. There is too much variation in the by-text slopes, in that they occupy both positive and negative logit space. This has an effect of “cancelling out” the overall effect.\(^32\) In other words, we cannot make general remarks about the effect of this predictor because different individuals seems to be doing markedly different things; this markedly different behaviour could therefore well be present in the abstract population for this predictor.

\(^{32}\)Separate analyses of the individual texts will not be carried out here, as idiolectal variation is not particular pertinent to the present study.
Figure 9.5: Individual by-text coefficients for GLMER rhythmic information on the participle/auxiliary boundary
CHAPTER 9. PHONOLOGY

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | 0.17700 | 0.20975 | 0.844 | 0.399 |
| PartAuxRhythm2clash | 0.02574 | 0.15011 | 0.172 | 0.864 |
| PartAuxRhythm2lapse | -0.01720 | 0.21795 | -0.079 | 0.937 |

Table 9.19: Final glmer for serialization ~ containing rhythmic information on the participle/auxiliary boundary

9.3.5.2 Auxiliary/Participle Rhythm

The overall distribution of this first rhythmic predictor is given in Table 9.20. An asymmetry between all of the factor levels is apparent, with clashes most strongly associating with the factus est order (69.56%) and lapses on the whole favouring est factus orders (57.89%). This apparent asymmetry is supported by a straightforward $\chi^2$-test of independence ($\chi^2 = 93.52$, $df = 2$, $p < 0.001$).

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>harmonious</td>
<td>435 (39.58)</td>
<td>664 (60.42)</td>
</tr>
<tr>
<td>clash</td>
<td>260 (30.44)</td>
<td>594 (69.56)</td>
</tr>
<tr>
<td>lapse</td>
<td>264 (57.89)</td>
<td>192 (42.11)</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
</tr>
</tbody>
</table>

Table 9.20: Relationship between auxiliary-participle rhythm and serialization

The four models to be constructed and compared are as follows. The first model includes the rhythmic predictor and varying intercepts for lemma and text:

```
auxpartrhythm.ri<-glmer(RealizationOfPAC~AuxPartRhythm2+
                         (1|LemmaParticiple)+(1|Text),pac,family=binomial)
```

The second model is the same as the first, with the addition of varying slopes for both lemma and text:

```
auxpartrhythm.rs.both<-glmer(RealizationOfPAC~AuxPartRhythm2+
                             (AuxPartRhythm2|LemmaParticiple)+(AuxPartRhythm2|Text),pac,family=binomial)
```

In the third model, the formula is the same as the first, but with the addition of varying slopes for lemma:

```
auxpartrhythm.rs.lemma<-glmer(RealizationOfPAC~AuxPartRhythm2+
                               (AuxPartRhythm2|LemmaParticiple)+(1|Text),pac,family=binomial)
```

The fourth model is also the same as the first, but this time with the addition of varying slopes for text:

```
auxpartrhythm.rs.text<-glmer(RealizationOfPAC~AuxPartRhythm2+
                            (1|LemmaParticiple)+(AuxPartRhythm2|Text),pac,family=binomial)
```

A comparison of the deviance and $AIC$ (along with other salient pieces of statistical information) is presented in Table 9.21. As for the previous predictor, the only two varying slope models that are significantly more informative than the more basic varying intercepts model are `auxpartrhythm.rs.both` ($\chi^2 = 24.433$, $df = 10$, $p = 0.00653$) and `auxpartrhythm.rs.text` ($\chi^2 = 22.904$, $df = 5$, $p = 0.0003522$). Furthermore, comparing the deviance of `auxpartrhythm.rs.both` with the nested models `auxpartrhythm.rs.text` and `auxpartrhythm.rs.lemma` in turn, we find that only `auxpartrhythm.rs.lemma` is significantly poorer ($\chi^2 = 18.708$, $df = 5$, $p = 0.002178$), meaning that varying slopes for lemma are redundant. As such, we adopt the model that is simultaneously the most informative but also the most statistically parsimonious, namely `auxpartrhythm.rs.text`. Note that this model has the lowest $AIC$ of all the models reported in Table 9.21.

A comparison of the selected glmer with the baseline vcm indicates that the inclusion of the predictor is a clear enhancement in reducing the amount of deviance in the response variable (cf. Table 9.22).
9.3. RHYTHMICAL OPTIMIZATION

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>auxpartrhythm.ri</td>
<td>5</td>
<td>3056.40</td>
<td>3085.33</td>
<td>-1523.20</td>
<td>3046.40</td>
</tr>
<tr>
<td>auxpartrhythm.rs.lemma</td>
<td>10</td>
<td>3060.67</td>
<td>3118.54</td>
<td>-1520.34</td>
<td>3040.67</td>
</tr>
<tr>
<td>auxpartrhythm.rs.text</td>
<td>10</td>
<td>3043.50</td>
<td>3101.37</td>
<td>-1511.75</td>
<td>3023.50</td>
</tr>
<tr>
<td>auxpartrhythm.rs.both</td>
<td>15</td>
<td>3051.97</td>
<td>3138.77</td>
<td>-1510.98</td>
<td>3021.97</td>
</tr>
</tbody>
</table>

Table 9.21: ANOVA statistics for various glmer s containing rhythmic information on the auxiliary/participle boundary

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.vcm</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>auxpartrhythm.rs.text</td>
<td>10</td>
<td>3043.50</td>
<td>3101.37</td>
<td>-1511.75</td>
<td>3023.50</td>
<td>109.91</td>
<td>7</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 9.22: ANOVA of baseline vcm and selected glmer containing rhythmic information on the auxiliary/participle boundary

We are now in a statistically valid position to consider the direction and strength of parameter estimates, which are given in Table 9.23. The estimate for intercept indicates the logit probability of finding a *factus est* outcome when the underlying rhythm in an *est factus* order is harmonious: it is 0.16 logits more likely, which equates to an odds ratio of 1.18. The other factor levels are compared to this baseline level. When the underlying rhythm in an *est factus* order results is a clash of strong syllables, e.g. *sánt dáti* [ssw], the logit probability of a *factus est* serialization increases by 0.47 logits (to 0.63); this is equivalent to an odds ratio of 1.60 compared to when the underlying rhythm would otherwise be harmonious. The absolute z-value (the estimate 0.4690 divided by the standard error 0.1378) is greater than the critical z-value for large sample sizes of 1.96, so we can have confidence in the parameter estimate: the exact p-value is very small at 0.0006, indicating that the likelihood of obtaining different results is less than 1 in 1000 different samples. As a further indication of the robustness of this parameter estimate, the confidence intervals lie above 0 in logits (0.20 to 0.74) and above 1 in odds (1.22 to 2.09). The result obtained here is evidently in the direction expected as per the Principle of Rhythmic Alternation. Moving down to the second parameter, when the underlying rhythm in an *est factus* order results in a lapse of strong syllables, e.g. *sínus ingrésst*, the logit probability of *factus est* reduces by −0.4897 logits to −0.3274; this is equivalent to an odds ratio of 0.6128102 as compared when the underlying rhythm would otherwise be harmonious. This parameter estimate is also significant at the p < 0.05 level of significance. In contrast to accentual clashes, lapses favour *est factus* serializations overwhelmingly, which appears to run counter to the Principle of Rhythmic Alternation. The theoretical importance of this will be discussed below.

|                  | Estimate | Std. Error | z value | Pr(> |z|) |
|------------------|----------|------------|---------|-------|-----|
| (Intercept)      | 0.1623   | 0.1949     | 0.833   | 0.404922 |
| AuxPartRhythm2clash | 0.4690   | 0.1378     | 3.403   | 0.000665 |
| AuxPartRhythm2lapse  | -0.4897 | 0.2345     | -2.089  | 0.036743 |

Table 9.23: Final glmer for serialization ∼ containing rhythmic information on the auxiliary/participle boundary

With this predictor, all the individual by-text slopes for the level clash are well above 0 in log odds (see Figure 9.6, with the strongest effect found in Nepos). For the level lapse, the majority lie in negative logit space, with two straddling around a logit value of 0 (*B.Afr. and B.Hisp*), and one (*Varro RR*) with an individual slope of 0.23, suggesting a more marked difference from the rest for this text.

9.3.6 Discussion

It was hypothesised first that *factus est* sequences would be dispreferred if the underlying rhythmical structure in a *factus est* order resulted in an accent clash or a lapse as opposed to a harmonious sequence
Figure 9.6: Individual by-text coefficients for GLMER rhythmic information on the auxiliary/participle boundary
in line with the Principle of Rhythmic Alternation and second that factus est serializations would be preferred if the underlying rhythmical structure in an est factus order resulted in in an accent clash or a lapse as opposed to a harmonious sequence. The data only partially support these hypotheses, with for instance underlying accentual clashes in an est factus order resulting in a higher probability of a factus est outcome than when the underlying sequence is perfectly harmonious.

However, the hypotheses are not supported on two accounts. First, underlying accentual lapses in an est factus order actually increase the probability of an est factus serialization, rather than promoting factus est outcomes. In this regard, it has been observed that lapses are less of a phonological violation than clashes. One possibility for Latin is that accentual lapses were perfectly acceptable.

Second, underlying accentual patterns in a factus est order have absolutely no main effect, as there is significant variability between texts (in some texts there is a positive effect, in other texts there is a negative effect). It is not immediately obvious why this should be so. It may be that rhythmical optimization has an effect only on certain types of phonological phrases, and crucially only within them and not between them; if factus est and est factus are differently phrased in terms of their prosodic (and maybe syntactic) structure, this would make sense.

There are several other reasons why some of the hypotheses were not borne out by the data. Again, it may be that written media are for the most part an inappropriate source to examine rhythmical effects, because of the considered and calculated nature of writing. One further possibility may be in how the variable has been operationalized. It may also be of interest to examine how the rhythmical sequences across the boundaries of the preverbal word and the PAC have an effect on serialization.

9.4 Segmental optimization

9.4.1 Introduction

We turn here to a related, but separate, phonological concept to the above, namely segmental optimization, that is how information about the segmental constraints of the language feed the syntax and determine the segmentally most optimal output.

In broad terms, segmental optimization is defined by the so-called Obligatory Contour Principle (ocp). The OCP is a language universal phonological constraint that asserts that adjacent identical segments are prohibited (McCarthy, 1986; Odden, 1986; Yip, 1988). For example, this constraint says that the adjacent sequence of voiceless velar plosives /k k/ would be avoided as would the adjacent sequence of vowels /a a/ because the segments are of identical articulatory make-up. Identical obstruents may also be avoided even if a syllabic nucleus intervenes and the two obstruents are not in immediate string adjacency, as happens in Arabic roots (McCarthy, 1986).

Research within the framework of probabilistic phonology has found that there is a gradience of phonotactic prohibition depending on how many and the type of phonological properties they share (cf. e.g. Pierrehumbert 1993; Frisch et al. 2004). For example, the sequence of segments /b g/ may be avoided because they share two articulatory features, namely [+voice] and [+plosive]: [b] is a voiced labial plosive, while [g] is a voiced velar plosive. Thus segmental similarity is of importance as well as segmental identity.

When the OCP is underlyingly violated, it often turns out that phonological processes ‘come to the rescue’ at the surface. For instance, in RP English, car alarm has an underlying representation /kA: disarm/ containing two vowels in hiatus. Because this hiatus is illicit according to the OCP, an intrusive-/r/ is inserted on the surface to optimize the phonological structure, as in [kA: r disarm].

---

33This was originally proposed by Leben (1973, 1980) for nonautosegmental phonology and extended to segmental phonology by McCarthy (1986).

34This constraint is sometimes known as horror aequi ‘bristling of the same’ in the usage-based literature; cf. Szmrecsanyi (2006: 39–40); Hinrichs & Szmrecsanyi (2007: 452).
We observe similar phonological strategies when the OCP is underlyingly violated in Latin. This has an effect on the synchronic and diachronic grammar. For example, when two vowels are adjacent (“vowel junction” in the terminology of Allen 1978: 78–82), the first vowel is elided: the string *profecta est is underlyingly represented as */pr6fEkta est/ and is realized as */pr6fEktext/. A historical example would be the degemination of EL caussa to CL causa (e.g. Sihler 1994: §232.2). A search of the LLT-A database, shows that even in Plautus (EL) caussa is underrepresented with 12 hits (6.97%) compared with causa which has 172 hits (93.02%). In Cicero (CL), by contrast, there is only a single example of caussa compared with 1933 hits for causa. Thus, the OCP has a clear impact on surface phonological realization, especially in the CL language. Another example of OCP in operation in surface Latin morphophonology concerns the adjectival suffix -alis in words such as *uenalis (e.g. Steriade 1987: 351–352). In stems that include a lateral, the allomorph -aris is used, as in *lunalis → lunaris. Here we see a clear dissimilatory effect, driven by the OCP.

9.4.2 Grammatical variability

9.4.2.1 General discussion

While segmental phonological constraints such as the OCP have an internal effect on the phonology, there has been far less discussion regarding the extent to which it can influence syntactic serializations and drive grammatical choices. I would imagine that this is because of the Chomskyan view that phonology has a limited role in syntax, as noted above. When we delve into the available literature, there are essentially two conflicting findings.

Jaeger (2006) investigates the use of optional that in complement clauses and relative clauses as a function of various phonological factors relating to the OCP. He hypothesizes that, if the effect of OCP is real in syntax, then that might be omitted if its inclusion would otherwise result in an identical or similar segmental sequence; similarly, that might be overtly realized if its omission would otherwise result in an identical or similar segmental sequence. Identity and similarity are defined and operationalized in terms of shared manner and place of articulation, but not voice information: thus, [b p] are treated as identical as they share manner of articulation (plosive) and place of articulation (labial), [b k] are treated as similar as they share manner of articulation (plosive) but differ in place of articulation (labial and velar, respectively), and [b s] are treated as different as they differ in manner of articulation (plosive and fricative, respectively) and differ in place of articulation (labial and alveolar, respectively). Through a series of logistic regression analyses, Jaeger shows that optional that is not used for segmental optimization either in complement clauses or in relative clauses. The results are either non-significant or in a direction counter to that predicted by a segmental optimization theory. Consequently, Jaeger (2006) concludes “The available evidence ... supports the Principle of Phonology-Free Syntax. Segmental optimization does not seem to affect (or at least not strongly affect) syntactic production”.

By contrast there are other studies which suggest that segmental phonology is important. Schütter (2003), for example, investigates the choice of the form of be in Chaucer’s language (late 14th century). In Chaucer’s language, be could alternate between a form that ended with a vowel be/bee and one that ended with a nasal consonant ben/been. Schütter (2003) examines the relationship between these two alternative forms and the initial phoneme of the following word. It is shown that, overall, while a following vowel-initial word selects be/bee in 3% of instances, a following consonant-initial word selects it in 49% of instances. Conversely, following vowel-initial words select ben/been in 97% of instances, whereas following consonant-initial words select it in 51% of instances. Thus, there is a strong asymmetry in the expected direction, whereby a CV/VC sequence is maintained rather than forcing a CC/VV clash. She also considers the glottal fricative h-. This selects ben/been in 91% of instances, but it is an

35I am assuming that the est here is non-clitic for the purpose of illustration. The representations are drawn from information in Allen (1978).
36Be that as it may, the use of orthographic to determine the surface phonology is of course problematic.
important separate category, as it significantly differs from the selection of following vowel-initial words. Compared to baseline choices of 35% and 65% for be/bee and ben/been respectively, analysis shows that hiatus is strongly dispreferred and consonant clusters are also dispreferred. Schlüter (2003) summarizes: “The explanation for this pattern of variation is now obvious, given the universal tendency to favour an ideal CV syllable structure wherever variants are available. On the one hand, the presence of -n before vocalic onsets promotes the avoidance of hiatuses, and on the other, the omission of -n before consonantal onsets serves the reduction of consonantal clusters”. As a consequence, Schlüter (2003: 101) resoundingly rejects the notion that phonology cannot drive syntax: “we can safely assume that the way language is phonologically implemented has repercussions on its grammar (both synchronically and diachronically). Therefore, a theory of grammar aiming at descriptive adequacy must have a means of making morphosyntactic choices dependent on the phonological form of its output”.

9.4.2.2 Segmental optimization in AVCs

With regard to AVC/PAC research, Muldowney (1937: 134), writing on St. Augustine, mentions that the est factus variant can be used to prevent hiatus or elision when (i) the participle ends in a vowel or a sequence containing a vowel plus an elidable consonant (e.g. -m) and (ii) the auxiliary begins with a vowel. In (35), the est factus variant is used to avoid a clash between the /-ə/ phoneme at the end of the participle and the /ɛ/- phoneme at the beginning of the auxiliary. Similarly, in (36), est relatum is chosen to prevent elision occurring in the sequence /-om e-/.

(35) quod violenter est passa
REL.N.SG.ACC violently be.3SG.PRS.IND suffer.PTCPL.F.SG.NOM
‘which she violently suffered’
(Aug. C. D. 1. 19, in Muldowney 1937: 134)

(36) quod in easdem litteras est relatum
REL.N.SG.ACC in same.F.PL.ACC letter.F.PL.ACC be.3SG.PRS.IND relate.PTCPL.N.SG.NOM
‘what has been related in the same letter’
(Aug. C. D. 21. 6, in Muldowney 1937: 134)

The avoidance of hiatus/elision as an effect on PAC alternation may only be a “tendency”, as Muldowney (1937: 134) mentions, because it fails to operate in the following examples, sourced from St. Augustine’s De Civitate Dei.

(37) et quia iram dei talem propterea passa
and because anger.F.SG.ACC god.M.SG.GEN such.F.SG.ACC moreover suffer.PTCPL.F.SG.NOM
est illa gens
be.3SG.PRS.IND DEM.F.SG.NOM race.F.SG.NOM
‘and because moreover that race suffered such anger of god’ (Aug. C. D. 18. 32)

(38) quia se facturum esse praedixit
because REFL.M.SG.ACC do.PTCPL.FUT.M.SG.ACC be.PRS.IND predcl.3SG.PRF.IND
‘because he predicted that we would do it’ (Aug. C. D. 21. 7)

One might note, however, that in these cases, had the PAC been serialized as est factus, it would have caused an elidable environment between the last preverbal word and the auxiliary, thus rendering the serialization switch as pointless with no phonological gain (assuming the relevance of this factor). For instance, although /-ə/ of passa and /ɛ/- of est are in hiatus/elidable, propterea est would also have provided an identical hiatus/elidable environment as passa est. In the second example, facturum is

37Kruschwitz (2004: 51, n. 114), discussing the position of esse in Republican inscriptions, has hinted at this factor as well. On the word order in hoc est factum monumentum Maerco Caicilio ‘this monument was made for Marcus Caecilius’ (CIL I 2 1202, 1–2), he remarks “Wurde est factum statt factum est geschrieben, um Elision bzw. Synaloephe zu vermeiden?”
elidable with esse, but the string se esse also produces an environment that could potentially result in hiatus/elision. Thus, a satisfactory investigation of elision/hiatus avoidance needs to take into account not only whether it can take place between the participle and the auxiliary (in that order), but also crucially whether it can take place between the end of the last preverbal word and the beginning of the auxiliary.

In summary, the views of Muldowney (1937) appear to support the notion of a syntax that is driven in part by segmental phonology, but clearly a detailed empirical investigation is needed to verify this.

9.4.3 Operationalization and variable profile

Four variables are configured with respect to OCP effects.

- First, I investigate the effect of the identity between the final segment of the participle and the initial segment of the auxiliary. One could hypothesize that if the segments are identical (e.g. passo est, where they are both vowels) or similar (e.g. passus fuit, where they are both fricatives) in a factus est serialization, then the alternative serialization might be used to repair and optimize the segmental phonology. This variable is termed PartAuxSegment.

- Second, I investigate the reverse of this, namely the identity between the final segment of the auxiliary and the initial segment of the participle. One could hypothesize that if the segments are identical (e.g. erg amatus, where they are both vowels) or similar (e.g. est conuulsus, where they are both plosives) in an est factus serialization, then the alternative serialization might be used to repair and optimize the segmental phonology. This variable is termed AuxPartSegment.

- Third, I investigate the effect of the identity between the final segment of the preverbal word and the initial segment of the participle. One could hypothesize that if the segments are identical (e.g. non natus est, where both are nasal dentals/alveolars) or similar (e.g. non reversus est, where both are dentals/alveolars) in the factus est serialization, then the alternative serialization might be used to repair and optimize the otherwise violated segmental sequence. This variable is termed PreverbalPartSegment.

- Fourth, I investigate the effect of the identity between the final segment of the preverbal word and the initial segment of the auxiliary. One could hypothesize that if the segments are identical (e.g. a Caesare est uocatus, where both are vowels) or similar (e.g. Caesar sit profectus, where both are coronal) in the est factus serialization, then the alternative serialization might be used by the speaker to repair and optimize the otherwise violated segmental phonology. This variable is termed PreverbalAuxSegment.

These four predictor variables are multiclass factors with a similar number of levels, so I will discuss their operationalization in tandem. The present operationalization draws in part on that by Jaeger (2006) and Frisch et al. (2004). Specifically, I identify \( k = 4 \) levels: identical, similar, different, and elidable. Like Jaeger (2006), the levels are distinguished in terms of manner and place of articulation. Using Allen (1978), the standard guide for Latin pronunciation, I distinguish seven manners of articulation (plosive, nasal, trill, lateral, fricative, approximant, and vowel) and eight places of articulation (bilabial, dental/alveolar, alveolar, velar, labio-velar, glottal, palatal, and vowel). Examples of Latin consonants are given in Table 9.24. (Note that I do not make reference to voicing or aspiration here.) Using this information, the factor levels are distinguished in the following ways:

- Two segments are regarded as identical if they share both manner and place of articulation.

- Two segments are regarded as similar if they share either manner or (at least one) place of articulation. Thus, the labio-dental fricative /f/ is similar to /b/ and /d/ in this operationalization, because /b/ is labial and /d/ is dental (or at least coronal in broad phonological terms).
9.4. SEGMENTAL OPTIMIZATION

- Two segments are regarded as **different** if they differ in both manner and place of articulation.
- Two segments are regarded as **elidable** if the first segment is a bilabial nasal /\textipa{m}/ and the second is any vowel (see the discussion on “vowel juncture” in Section 9.4.1 above). Note that the level elidable applies only to immediately adjacent segments, and not the variables that are designed to tap long-distance OCP violations.

<table>
<thead>
<tr>
<th>Grapheme</th>
<th>Example</th>
<th>Phoneme</th>
<th>Manner of Articulation</th>
<th>Place of Articulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>profectus ‘departed’</td>
<td>/p/</td>
<td>plosive</td>
<td>bilabial</td>
</tr>
<tr>
<td>ph</td>
<td>philosophatus ‘philosophized’</td>
<td>/p\textipa{b}/</td>
<td>plosive</td>
<td>bilabial</td>
</tr>
<tr>
<td>b</td>
<td>bellatum ‘waged war’</td>
<td>/b/</td>
<td>plosive</td>
<td>bilabial</td>
</tr>
<tr>
<td>t</td>
<td>traditus ‘handed over’</td>
<td>/\textipa{t}, \textipa{t}/</td>
<td>plosive</td>
<td>dental/alveolar</td>
</tr>
<tr>
<td>d</td>
<td>datus ‘given’</td>
<td>/\textipa{d}, \textipa{d}/</td>
<td>plosive</td>
<td>dental/alveolar</td>
</tr>
<tr>
<td>c</td>
<td>conditus ‘founded’</td>
<td>/k/</td>
<td>plosive</td>
<td>velar</td>
</tr>
<tr>
<td>g</td>
<td>gavisus ‘rejoiced’</td>
<td>/g/</td>
<td>plosive</td>
<td>velar</td>
</tr>
<tr>
<td>qu</td>
<td>quaesitus ‘searched for’</td>
<td>/k\textipa{w}/</td>
<td>plosive</td>
<td>labio-velar</td>
</tr>
<tr>
<td>n</td>
<td>natus ‘born’</td>
<td>/\textipa{n}, \textipa{n}/</td>
<td>nasal</td>
<td>dental/alveolar</td>
</tr>
<tr>
<td>m</td>
<td>moratus ‘delayed’</td>
<td>/\textipa{m}/</td>
<td>nasal</td>
<td>bilabial</td>
</tr>
<tr>
<td>r</td>
<td>ratus ‘thought, ratified’</td>
<td>/\textipa{r}, \textipa{r}/</td>
<td>trill</td>
<td>dental/alveolar</td>
</tr>
<tr>
<td>l</td>
<td>relatus ‘reported’</td>
<td>/\textipa{l}, \textipa{l}/</td>
<td>lateral</td>
<td>dental/alveolar</td>
</tr>
<tr>
<td>f</td>
<td>factus ‘done’</td>
<td>/\textipa{f}/</td>
<td>fricative</td>
<td>labio-dental</td>
</tr>
<tr>
<td>s</td>
<td>situs ‘situated’</td>
<td>/\textipa{s}/</td>
<td>fricative</td>
<td>alveolar</td>
</tr>
<tr>
<td>h</td>
<td>horatus ‘encouraged’</td>
<td>/\textipa{h}/</td>
<td>fricative</td>
<td>glottal</td>
</tr>
<tr>
<td>i</td>
<td>iactus ‘thrown’</td>
<td>/\textipa{j}/</td>
<td>approximant</td>
<td>palatal</td>
</tr>
<tr>
<td>u</td>
<td>uitus ‘conquered’</td>
<td>/\textipa{w}/</td>
<td>approximant</td>
<td>labio-velar</td>
</tr>
</tbody>
</table>

Table 9.24: Latin consonantal phonemes by manner and place of articulation (information drawn from Allen 1978)

9.4.4 Statistical modelling

9.4.4.1 Final segment of the participle and initial segment of the auxiliary

Table 9.25 presents the overall distribution of this variable with respect to PAC serialization choice. The variable levels within this crosstabulation exhibit some fluctuations about the mean, but they are to some extent difficult to interpret. First, segments that are different (e.g. datus est, which consists of an alveolar fricative followed by a vowel) associate overwhelmingly with the factus est serialization (61.53% of instances); this is easily interpretable, as it is in the direction predicted by the OCP, wherein the factus est serialization produces an optimal segmental sequence. However, while identical segments slightly disfavour factus est, it appears that similar segments disfavour this serialization the strongest (at 49.15%). If the OCP was being conformed to reliably, then we would expect identical segments (such as passa est or interfectus sit) to disprefer factus est the most strongly, because they result in suboptimal segmental sequences. Finally, it is unclear why elidable segments pattern very much like different segments with regard to their distribution — if anything, one should expect them to be more like identical segments on phonological grounds (in that a vowel-vowel and bilabial nasal-vowel sequence would typically result in the first segment being deleted). The distribution is only statistically marginally significant on a \( \chi^2 \)-test (\( \chi^2 = 6.75, df = 3, p = 0.08 \)). A discussion based merely on percentages and a bivariate \( \chi^2 \)-test could well be misleading, however. Consequently, we need to inspect the significance of this distribution in more detail via the appropriate statistical models.

Let us therefore begin constructing a model for statistical interpretation. As usual four models are constructed and compared. First, a varying intercepts model is constructed:

\[
\text{partaux.ri}<\text{glmer(RealizationOfPAC\textasciitilde(1|LemmaParticiple)+(1|Text)+PartAuxSegment,pac,family=binomial)}
\]
Chapter 9. Phonology

Table 9.25: Relationship between the participle-auxiliary segmental phonology and serialization

<table>
<thead>
<tr>
<th>Serialization</th>
<th>factus est (%)</th>
<th>factus est (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>identical</td>
<td>244 (42.96)</td>
<td>324 (57.04)</td>
</tr>
<tr>
<td>similar</td>
<td>30 (50.85)</td>
<td>29 (49.15)</td>
</tr>
<tr>
<td>different</td>
<td>464 (38.47)</td>
<td>742 (61.53)</td>
</tr>
<tr>
<td>elidable</td>
<td>221 (38.37)</td>
<td>355 (61.63)</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
</tr>
</tbody>
</table>

Table 9.26: ANOVA statistics for various glmerS containing final participle/initial auxiliary segmental information

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>partaux.ri</td>
<td>6</td>
<td>3140.23</td>
<td>3174.95</td>
<td>-1564.12</td>
<td>3128.23</td>
</tr>
<tr>
<td>partaux.lemma</td>
<td>15</td>
<td>3141.05</td>
<td>3227.86</td>
<td>-1555.53</td>
<td>3111.05</td>
</tr>
<tr>
<td>partaux.rs.text</td>
<td>15</td>
<td>3127.79</td>
<td>3214.59</td>
<td>-1548.89</td>
<td>3097.79</td>
</tr>
<tr>
<td>partaux.rs.both</td>
<td>24</td>
<td>3134.08</td>
<td>3272.96</td>
<td>-1543.04</td>
<td>3086.08</td>
</tr>
</tbody>
</table>

Table 9.27: ANOVA output comparing the four models is reported in Table 9.26. Compared to the varying intercepts model partaux.ri, the three more complex models are substantial improvements statistically, in that they result in significantly reduced deviances and their AIC values are closer to zero. Comparing the most complex model partaux.rs.both with the nested models — partaux.lemma and partaux.rs.text —, it is apparent that, while the model that includes only varying slopes for lemma is significantly poorer than the one that includes varying slopes for both lemma and text, the model that includes only varying slopes for text is not significantly poorer than the one that includes varying slopes for both lemma and text. This indicates that there is significant variability with respect to differences between texts, but not between verb lemmas, for this predictor. We therefore choose for further appraisal the model that is the most informative but at the same time the most parsimonious, namely partaux.rs.text.

This model is then compared with the baseline vcm. The ANOVA output in Table 9.27 demonstrates that the inclusion of the participle-auxiliary segmental information is a significant improvement.

With the final model for this predictor now established, let us turn to the parameters, reported in Table 9.28. As usual, the response variable is in terms of the logit probability of factus est (= 1), and the predictor variable is dummy coded with the level different established as the baseline (= 0 throughout).

Table 9.28: Parameters for the final model partaux.rs.text

<table>
<thead>
<tr>
<th></th>
<th>different</th>
<th>similar</th>
<th>elidable</th>
</tr>
</thead>
<tbody>
<tr>
<td>different</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

128
### 9.4. SEGMENTAL OPTIMIZATION

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>glmer.vcm.both</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>partaux.rs.text</td>
<td>15</td>
<td>3127.79</td>
<td>3214.59</td>
<td>-1548.89</td>
<td>3097.79</td>
<td>35.62</td>
<td>12</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Table 9.27: ANOVA of baseline VCM and selected GLMER containing final participle/initial auxiliary segmental information

<table>
<thead>
<tr>
<th></th>
<th>identical</th>
<th>similar</th>
<th>elidable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Given that none of the coefficients is significant, it appears that the segmental junction between the final segment of the participle and the initial segment of the auxiliary plays no role in serialization choice.

Table 9.28: Final GLMER for serialization \(\sim\) final participle/initial auxiliary segmental information

|                          | Estimate | Std. Error | z value | Pr(>|z|) |
|--------------------------|----------|------------|---------|---------|
| (Intercept)              | 0.16398  | 0.22434    | 0.731   | 0.465   |
| PartAuxSegmentidentical  | -0.03693 | 0.16608    | -0.222  | 0.824   |
| PartAuxSegmentsimilar    | -0.07345 | 0.57863    | -0.127  | 0.899   |
| PartAuxSegmentelidable   | 0.18618  | 0.19581    | 0.951   | 0.342   |

One might wonder why the model is significant as a whole, but the parameter estimates are not. For instance, with other variables we have seen that, while the varying slope component can be influential, the parameter estimates are still significant. This is not the case here. The reason for this can be discerned from the plots in Figure 9.7. Each of the plots displays the individual slopes for the texts, for (i) identical segments compared to different segments, (ii) similar segments compared to different segments, and (iii) elidable segments compared to different segments. The plots show that for each of the three levels, there is too much variation between texts, with slopes both in negative and positive space. This is important. On the one hand, varying slopes can be influential and parameters significant if there is a lot of variability in the slopes and the majority are in either positive or negative space. On the other hand, varying slopes can be influential and parameters not significant if there is a lot of variability in the slopes, but they are scattered fairly evenly in positive and negative space. It is that latter condition we see afflicting the present predictor.

#### 9.4.4.2 Final segment of the auxiliary and initial segment of the participle

Elidable segments, which have only a few observations \((n = 11)\) for this predictor, were included within the level for identical segments, because of their similar phonological effects (in e.g. poetry). Finally, the factor levels identical and similar were collapsed as similar because they are not statistically significantly different according to a pairwise test of proportions with Holm’s adjustment \((p = 0.77)\). So we now have a binary \(k = 2\) predictor, distinguishing similar vs. different. Table 9.29 shows the overall distribution for this variable. Impressionistically, there is an effect whereby the OCP is being conformed to: if the resulting sequence in the est factus serialization is optimal, i.e. the segments are different, there is less of a need to revert to the factus est serialization than when the resulting sequence in the est factus serialization is suboptimal, i.e. the segments are similar (including identical and elidable ones). The distribution is statistically significant on a \(\chi^2\)-test \((\chi^2 = 23.5732, df = 1, p = 1.203e - 06)\).

Let us turn to modelling this predictor in multilevel models that use Laplace approximation. Again, four models are constructed and compared. First, a varying intercepts model for both lemma and text:

```r
auxpart.ri<-glmer(RealizationOfPAC~AuxPartSegment+
                   (1|LemmaParticiple)+(1|Text),pac,family=binomial)
```

38See the discussion on “vowel-juncture” above. Statistically, however, it would not have mattered if elidable observations had been merged with any of the other levels, or had been omitted altogether.
Figure 9.7: Individual by-text coefficients for GLMER containing final participle/initial auxiliary segmental information
9.4. SEGMENTAL OPTIMIZATION

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>different</td>
<td>465 (45.50)</td>
<td>557 (54.50)</td>
</tr>
<tr>
<td>similar</td>
<td>494 (35.62)</td>
<td>893 (64.38)</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
</tr>
</tbody>
</table>

Table 9.29: Relationship between the auxiliary-participle segmental phonology and serialization

In the second model, varying slopes for both lemma and text are included as well:

```r
auxpart.rs.both<-glmer(RealizationOfPAC~AuxPartSegment+(AuxPartSegment|LemmaParticiple)+(AuxPartSegment|Text),pac,family=binomial)
```

In the third model, varying slopes are included for lemma but not for text:

```r
auxpart.rs.lemma<-glmer(RealizationOfPAC~AuxPartSegment+(1|Text),pac,family=binomial)
```

And in the fourth model, varying slopes are included for text, but not for lemma:

```r
auxpart.rs.text<-glmer(RealizationOfPAC~AuxPartSegment+(1|LemmaParticiple)+(AuxPartSegment|Text),pac,family=binomial)
```

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>auxpart.ri</td>
<td>4</td>
<td>3113.17</td>
<td>3136.32</td>
<td>-1552.58</td>
</tr>
<tr>
<td>auxpart.rs.lemma</td>
<td>6</td>
<td>3109.16</td>
<td>3143.88</td>
<td>-1548.58</td>
</tr>
<tr>
<td>auxpart.rs.text</td>
<td>6</td>
<td>3115.31</td>
<td>3150.03</td>
<td>-1551.65</td>
</tr>
<tr>
<td>auxpart.rs.both</td>
<td>8</td>
<td>3111.82</td>
<td>3158.12</td>
<td>-1547.91</td>
</tr>
</tbody>
</table>

Table 9.30: ANOVA statistics for auxiliary-participle segments

ANOVA statistics including deviance and AIC values are reported in Table 9.30. Compared to the least complex model `auxpart.ri`, it is only `auxpart.rs.lemma` that is statistically more firmly informative whilst being parsimonious at the same time. Clearly, text-based differences are of no account for this particular predictor, but lemmatic ones are. Compared to the baseline `vcm`, the model `auxpart.rs.lemma` is superior (cf. Table 9.31), so we will use it to inspect the parameters.

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.vcm</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>auxpart.rs.lemma</td>
<td>6</td>
<td>3109.16</td>
<td>3143.88</td>
<td>-1548.58</td>
<td>3097.16</td>
<td>36.25</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 9.31: ANOVA of baseline `vcm` and selected `glmer` containing final auxiliary/initial participle segmental information

The parameters for the resulting model are reported in Table 9.32. When the sequence is of a similar segmental nature, e.g. *es ferendus* where there is a sequence of fricative segments /s f-/`, the logit probability of a factus est outcome increases by 0.51. This is the equivalent of an odds ratio of 1.67. The 95% confidence intervals associated with this parameter estimate do not include 0 in logits (they range from 0.30 to 0.72) or 0 in odds (1.35 to 2.06), indicating its statistical significance. A further indication of its statistical significance is the fact that the absolute z-value (the estimate 0.51 divided by its standard error 0.1) is greater than 1.96, the critical z-value for large sample sizes. The effect is in line with that predicted by the `ocp`.

With the previous predictor, it was seen that text-based slopes were too great to result in statistically robust parameter estimates, because the individual by-text coefficients were both in negative and positive logit space. This is not the case with the lemma slopes, by contrast. Although there is variability, the
Table 9.32: Final glmer for serialization ~ final auxiliary/initial participle segmental information

|                         | Estimate | Std. Error | z value | Pr(>|z|) |
|-------------------------|----------|------------|---------|----------|
| (Intercept)             | -0.1283  | 0.1758     | -0.730  | 0.465    |
| AuxPartSegmentsimilar   | 0.5012   | 0.1090     | 4.679   | 2.88e-06 |

Individual by-lemma coefficients: similar

Figure 9.8: Individual by-lemma coefficients for glmer containing final auxiliary/initial participle segmental information

The majority of the individual by-lemma coefficients (589 out of 604) are in positive logit space, as displayed on the plot in Figure 9.8. The lemmas that have the polar opposite effect to the majority are: excipio ‘take out’, paro ‘prepare’, sum ‘be’, uideo ‘see’, audeo ‘dare’, inuenio ‘find’, uenio ‘come’, arbitror ‘think’, recipio ‘receive’, deligio2 ‘tie up’, hortor ‘urge on’, nosco ‘know’, educo1 ‘train’, orior ‘arise’, and insero ‘plant’.

9.4.4.3 Final segment of the preverbal word and initial segment of the participle

The previous two predictors concerned segmental effects within the PAC. For the next two predictors we shall examine segmental effects at the PAC’s left edge, beginning with the boundary between the preverbal word and the participle. For the readers’ interest, Table 9.33 shows the overall distribution of this variable cross-classified for the serialization of the PAC, which is only marginally significant on a χ²-test ($\chi^2 = 6.62$, $df = 3$, $p = 0.09$).

The first model includes varying intercepts for both lemma and text:

```r
prepart.r<-glmer(RealizationOfPAC~PreverbalPartSegment+(1|LemmaParticiple)+(1|Text), pac, family=binomial)
```
9.4. SEGMENTAL OPTIMIZATION

Table 9.33: Relationship between the preverbal/participle segmental phonology and serialization

<table>
<thead>
<tr>
<th></th>
<th>Serialization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est factus (%)</td>
</tr>
<tr>
<td>different</td>
<td>616 (38.72)</td>
</tr>
<tr>
<td>identical</td>
<td>136 (43.04)</td>
</tr>
<tr>
<td>similar</td>
<td>161 (39.17)</td>
</tr>
<tr>
<td>elidable</td>
<td>46 (50.55)</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>959</td>
</tr>
</tbody>
</table>

The second model is the same as the first model, with the addition of varying slopes for both lemma and text:

prepart.rs.both<-glmer(RealizationOfPAC~PreverbalPartSegment+(PreverbalPartSegment|LemmaParticiple)+(PreverbalPartSegment|Text),pac,family=binomial)

The third model includes varying slopes for lemma but not text:

prepart.rs.lemma<-glmer(RealizationOfPAC~PreverbalPartSegment+(PreverbalPartSegment|LemmaParticiple)+(1|Text),pac,family=binomial)

The fourth model includes varying slopes for text but not lemma:

prepart.rs.text<-glmer(RealizationOfPAC~PreverbalPartSegment+(1|LemmaParticiple)+(PreverbalPartSegment|Text),pac,family=binomial)

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>prepart.ri</td>
<td>6</td>
<td>3141.18</td>
<td>3175.90</td>
<td>-1564.59</td>
<td>3129.18</td>
</tr>
<tr>
<td>prepart.lemma</td>
<td>15</td>
<td>3150.64</td>
<td>3237.45</td>
<td>-1560.32</td>
<td>3120.64</td>
</tr>
<tr>
<td>prepart.rs.text</td>
<td>15</td>
<td>3158.10</td>
<td>3244.90</td>
<td>-1564.05</td>
<td>3128.10</td>
</tr>
<tr>
<td>prepart.rs.both</td>
<td>24</td>
<td>3167.15</td>
<td>3306.03</td>
<td>-1559.57</td>
<td>3119.15</td>
</tr>
</tbody>
</table>

Table 9.34: ANOVA statistics for preverbal-participle segments

The ANOVA output in Table 9.34 shows that none of the more complex models are significantly more informative than the simpler model, prepart.ri. For example, prepart.rs.both has a lower deviance of 3119.2, but the reduction in deviance of 10.03 (= 3129.18 – 3119.15) is not statistically significant on 18 (= 24 – 6) degrees of freedom. Notice further that the AIC of the simplest model is closer to zero, indicating that it should be preferred. We will therefore take up this model for further appraisal.

The next step is to compare this model with the baseline vcm in order to determine whether including the predictor is to any statistically meaningful advantage. The ANOVA of these two models is reported in Table 9.35. In contrast to the models established for many previous predictor variables considered thus far, this shows that the model including the predictor is not significantly better than the baseline vcm which contains merely varying intercepts for verb lemma and text. This is because the reduction in deviance of 4.23 (= 3133.41 – 3129.18) is not significant (p = 0.24) on 3 (= 6 – 3) degrees of freedom. As such, the baseline vcm is the final model for this predictor variable, and there are no parameters to explain.39 In short, the segmental properties of phonological boundary across the word on the pac’s left edge and the participle play no role in serialization choice.

39Actually, if the summary of this model is called in r, it turns out that one of the parameters, namely elidable vs. different, is statistically significant (β = −0.4668, SE = 0.2320, z = −2.012, p = 0.0442). However, one should not interpret such a model, because it is not significant as a whole. This illustrates the importance of comparing fuller and nested models with deviance, AICs, and other such measures before interpreting.
CHAPTER 9. PHONOLOGY

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.vcm</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prepart.ri</td>
<td>6</td>
<td>3141.18</td>
<td>3175.90</td>
<td>-1564.59</td>
<td>3129.18</td>
<td>4.23</td>
<td>0.2380</td>
</tr>
</tbody>
</table>

Table 9.35: ANOVA of baseline VCM and selected GLMER containing final preverbal/initial participle segmental information

9.4.4.4 Final segment of the preverbal word and initial segment of the auxiliary

Table 9.36 shows the overall distribution of this predictor variable cross-classified for the serialization of the PAC. An overall $\chi^2$-test reveals this distribution to be statistically significant ($\chi^2 = 18.08$, $df = 3$, $p < 0.001$); however, it is still important to model the effect of this variable appropriately, which we turn to now.

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>different</td>
<td>485 (43.77)</td>
<td>623 (56.23)</td>
</tr>
<tr>
<td>identical</td>
<td>308 (34.92)</td>
<td>574 (65.08)</td>
</tr>
<tr>
<td>similar</td>
<td>49 (45.37)</td>
<td>59 (54.63)</td>
</tr>
<tr>
<td>elidable</td>
<td>117 (37.62)</td>
<td>194 (62.38)</td>
</tr>
</tbody>
</table>

Table 9.36: Relationship between the preverbal-auxiliary segmental phonology and serialization

The first model includes varying intercepts for both lemma and text:

```r
preaux.ri <- glmer(RealizationOfPAC ~ PreverbalAuxSegment + (1 | LemmaParticiple) + (1 | Text), pac, family = binomial)
```

The second model is the same as the first model, with the addition of varying slopes for both lemma and text:

```r
preaux.rs.both <- glmer(RealizationOfPAC ~ PreverbalAuxSegment + (PreverbalAuxSegment | LemmaParticiple) + (PreverbalAuxSegment | Text), pac, family = binomial)
```

The third model includes varying slopes for lemma but not text:

```r
preaux.rs.lemma <- glmer(RealizationOfPAC ~ PreverbalAuxSegment + (PreverbalAuxSegment | LemmaParticiple) + (1 | Text), pac, family = binomial)
```

The fourth model includes varying slopes for text but not lemma:

```r
preaux.rs.text <- glmer(RealizationOfPAC ~ PreverbalAuxSegment + (1 | LemmaParticiple) + (PreverbalAuxSegment | Text), pac, family = binomial)
```

We now compare the models in terms of their deviances and AICs. To facilitate this end, the relevant ANOVA output is reported in Table 9.37. This shows that only one of the more complex models, namely `preaux.rs.text`, is significantly more informative than the varying intercepts model `preaux.ri`; `preaux.rs.lemma` and `preaux.rs.both`, by contrast, contain superfluous information. Lemmatic differences in the slopes are not important, but textual ones are.

Finally, it is worth comparing `preaux.rs.text` to the baseline VCM to determine whether the inclusion of the predictor is important at all, or whether we can just get away with the simplest model permissible. From Table 9.38, it appears that the addition of the predictor is essential in enhancing the statistical power of the model. We will therefore keep it in, and inspect the parameters. The resulting model (Table 9.39) shows that two of the parameters are significantly different from the baseline category of different. Specifically, if the segmental properties of the phonological boundary between the left-edge
9.4. SEGMENTAL OPTIMIZATION

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>preaux.ri</td>
<td>6</td>
<td>3128.73</td>
<td>3163.45</td>
<td>-1558.37</td>
<td>3116.73</td>
</tr>
<tr>
<td>preaux.rs.lemma</td>
<td>15</td>
<td>3145.45</td>
<td>3232.25</td>
<td>-1557.72</td>
<td>3115.45</td>
</tr>
<tr>
<td>preaux.rs.text</td>
<td>15</td>
<td>3127.95</td>
<td>3214.75</td>
<td>-1548.97</td>
<td>3097.95</td>
</tr>
<tr>
<td>preaux.rs.both</td>
<td>24</td>
<td>3144.49</td>
<td>3283.38</td>
<td>-1548.24</td>
<td>3096.49</td>
</tr>
</tbody>
</table>

Table 9.37: ANOVA statistics for preverbal-auxiliary segments

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chiq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chiq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.vcm</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>preaux.rs.text</td>
<td>15</td>
<td>3127.95</td>
<td>3214.75</td>
<td>-1548.97</td>
<td>3097.95</td>
<td>35.46</td>
<td>12</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Table 9.38: ANOVA of baseline VCM and selected GLMER containing final preverbal/initial auxiliary segmental information

word and the initial phoneme of the auxiliary are identical, then the logit probability of \textit{factus est} increases by 0.51 which is equivalent to an odds ratio of 1.72. The confidence intervals associated with this estimate range from 0.24 to 0.84 in logits; since they do not include zero, the parameter can be considered statistically significant. The extent to which it is, can be assessed by dividing the estimate 0.54 by its standard error of 0.15 to give a \textit{z-value} of 3.51. This value with the relevant degrees of freedom, gives a \textit{p-value} of < 0.001, indicating that the likelihood of obtaining different results is less than 1 in 1000. Similarly, for elidable segments, the logit probability of \textit{factus est} increases by 0.46 which is equivalent to an odds ratio of 1.58. The confidence intervals associated with this estimate are wider than the previous one, ranging from a logit of 0.05 to 0.87. Again, this parameter estimate is significant, as \textit{z} = 2.197 = 0.02801 > 1.96, indicating that the likelihood of obtaining different results is about twice in 100 samples.

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | -0.07222 | 0.21328 | -0.339 | 0.73491 |
| PreverbalAuxSegmentidentical | 0.54185 | 0.15416 | 3.515 | 0.00044 |
| PreverbalAuxSegmentsimilar | -0.12599 | 0.29245 | -0.431 | 0.66662 |
| PreverbalAuxSegmentelidable | 0.45830 | 0.20858 | 2.197 | 0.02801 |

Table 9.39: Final GLMER for serialization \sim final preverbal/initial auxiliary segmental information

The differences in by-text slopes are worth discussing briefly, and are plotted in Figure 9.9 for each of the three levels (identical, similar, and elidable) compared to the baseline. As this indicates, there is too much variability with respect to the level similar for anything meaningful to be said about it (the individual slopes are situated in both negative and positive logit space). However, for identical and elidable sequences, the effects are all positive with the exception of a single text, namely Cicero’s \textit{Philippics}. With regard to this predictor Cicero has set himself out from the pack.

9.4.5 Discussion

The preceding sections have been rather data and statistics dense, so we will turn here to an omnibus discussion of the regression models. To recapitulate, there are two influential predictors, namely \textit{AuxPartSegment} and \textit{PreverbalAuxSegment}, and two non-influential predictors, namely \textit{PartAuxSegment} and \textit{PreverbalPartSegment}. Specifically, if the final segment of the auxiliary and the initial segment of the participle are similar (or identical or elidable), then a \textit{factus est} serialization is more likely, because an \textit{est factus} serialization would result in a violation of the OCP: a good example of this is \textit{traditus est}, where the alternative serialization \textit{est traditus} would put two identical segments /-t t-/ in sequence, thereby violating the OCP. Similarly, if the final segment of the preverbal word and the initial segment of the auxiliary are identical or elidable, then the \textit{factus est} serialization is adopted to prevent an OCP violation: to exemplify, in \textit{naue captus est}, the alternative serialization \textit{naue est captus
Figure 9.9: Individual by-text coefficients for GLMER containing final preverbal/initial auxiliary segmental information
would put two identical segments in adjacency /-e e-/ resulting in a hiatus or in a necessity to elide the first. However, for the junction between the participle’s final syllable and the auxiliary’s initial syllable and for the junction between the preverbal word’s final syllable and the participle’s initial syllable, no such effects can be discerned.

In other words, the OCP appears to have an effect when the resulting serialization would be est factus. The results exactly mirror those found for the rhythmical optimization variable. Why should this be? One possibility concerns the prosodic and syntactic structure of the PAC unit. There is evidence that elements that share the same prosodic phrase are much more sensitive to phonological reduction, assimilation, etc. than those across higher phrase types (such as between phonological phrases and intonational phrases). If we accept that PACs in the factus est order are more of a tightly knit prosodic and syntactic ‘unit’ than those in est factus orders, it may be that OCP violations when the serialization would be factus est are remedied not through word order permutation but more straightforwardly through phonological level effects, such as elision, segmental reduction (such as epenthesis), and liaison, etc.

Theoretically, it is thus apparent that there is some evidence for conformity to OCP in syntax from the present analysis, and therefore the involvement of segmental phonology in syntax, which runs counter to the claims of Chomsky (1995) and the notion of a Phonology Free Syntax. One further piece of evidence from the present data analysis that a phonology free syntax is not supported. We should also note that it is evidently not categorical, but probabilistic.

However, to corroborate this bivariate evidence, we need to see whether these variables are still non-significant and do not participate in interaction effects in our multivariate modelling in Part V.

9.5.1 Background

Latin prose texts, particularly those which are supposedly “artistic” and oratorical in nature, are claimed to exhibit recurrent metrical patterns at the end of clauses, sentences, and prosodic phrases. These patterns are usually known as clausulae. Some metrical patterns, such as the double cretic [LHL|LHL] and cretic-spondee/trochee [LHL|HH], are favoured, whereas others, such as the dactyl-spondee “heroic” ending [HLL|HH], are largely avoided because, it is claimed, they evoke epic poetry, which was composed on a dactylic base. In this section we will therefore explore the extent to which metrical patterns in prose exert an influence on on the internal serialization of PACs.

9.5.2 Previously identified grammatical effects

Skutsch (1912) claims that clausulae preferences of an author may well have constituted a decisive factor on the choice of PAC variant. Similar remarks are to be found scattered in philological commentaries, including, but not limited to, those of Gotoff (1979, 1993), Dyck (1996, 2003, 2008, 2010), and works on specific authors, e.g. Muldowney (1937) on St. Augustine, Feix (1934) on Petronius, and Bloomer (1992) on Valerius Maximus. For example, Muldowney (1937: 134) claims that “[a] desire to secure a favorite clausula seems to have influenced the position of the copula in many instances”. Similarly, Bloomer (1992: 247) notes “Valerius does rearrange word order to achieve these clausulae as the placement of forms of esse demonstrates: Punica classis esset oppressa (2. 8. 2); leges a se esse servatas (3. 7. 1d)”; here, we have esset oppressa and esse servatas instead of oppressa esset and servatas esse, respectively, thus avoiding a long sequence of heavy syllables. Finally, Dyck asserts (2003: 205) “[where] there are

40 Zielinski 1904 is a ground-breaking study on this topic. See also Aili (1979) and Hutchinson (1995), who summarize and extend his analysis. Some scholars prefer to talk in terms of “rhythmical prose” rather than “metrical prose”, but I cannot see why the distinction is anything but terminological. In any case, I reserve the term “rhythm” for a different group of variables.

41 Bloomer (1992) appears to draw on an early dissertation by Muench (1909) for this information.
auxiliaries of periphrastic verbs, some manipulation of the expected order may be needed to achieve the right sonority”.

Be that as it may, the extent to which clausulae affect PAC serialization is not entirely clear-cut in my view, though. I mention here some examples from Dyck’s commentaries that raise questions. At Cic. Cat. 1. 4. 5 est permissā rēs públicā, he claims (2008: ad loc.) that the est factus variant is used to create a double cretic, which according to his list of Cicero’s favourite metrical shapes is Cicero’s third favourite (2008: 210). But had Cicero used the factus est variant permiss(a) est rēs públicā, a molossus-cretic sequence would have been produced, which is even more preferred than the double cretic (2008: ibid.).

Another case is that mentioned at Cic. Cat. 1. 33. 4 a Romul(o) es cōns(t)ūt(a) where the est factus variant is alleged to have been brought into service to effect a ditrochaic (i.e. trochee-trochee/spondee) clausula. Yet the factus est variant a Romulo cōns(t)ūt(a) ēs would have produced a clausula of an identical type. Consequently, Dyck’s qualitative claims — and possibly those of other scholars — do not appear to stand up to even a brief examination.42

It is important that we probe the effect of this factor, however, first because we want to understand what conditions the serialization choice for PACs but also because textual critics often use clausulae to help decide on word order variants in manuscripts. If it turns out to be uninfluential — and even it is not categorical —, textual critics may have to revise their methods substantially.

9.5.3 Variable profile and operationalization

The operationalization of any clausulae-related variable is not as straightforward as it might appear. There are problems annotating for whether a particular serialization was used to exploit or avoid particular metrical patterns, mainly because we do not know all the rules governing clausulae found in prose. Some of the re-arranged examples given above look metrically possible to us, but to what extent would they have been acceptable to the native ear? In spite of these problems, we must still attempt to code information about clausulae in some way, as it may well account for some of the variance in the response variable.43

The following procedure was used to determine the clausulae structure of PAC sequences.

We cannot simply scan each observation for its metrical structure. First, this would involve incorporating information about the response directly into the predictor. For instance, prōfectūs ērāt and ērāt prōfectūs have inherently different scansions. Hence, we would be using the word order to identify the scansion, and in turn then using the scansion to identify the word order. The result is circular, and we would be predicting — as it were — what we already know. Second, it is not obvious that any effect would be obtained anyway, particularly if a writer always strove to use similar metrical patterns. For instance, if a writer always used whichever order produced a double-trochee sequence, then each observation would be tagged as a double-trochee and there would be absolutely no variation to analyze. In order to assess the impact of clausulae, we therefore need to “normalize” the word order before scanning it. We are interested, moreover, in predicting which PAC serialization would be chosen if a string had a particular metrical sequence when scanned in a normalized order. A variable was therefore created utilizing R scripts out of convenience variables, whereby the participle form and the auxiliary form were combined in that order with a period separator:

\[
pac\text{\$PartAuxMetrics}=\text{paste}(pac\text{\$Participle},pac\text{\$Auxiliary},sep=\text{\".\")}
\]

\[
pac\text{\$PartAuxMetrics}=\text{as.factor}(pac\text{\$PartAuxMetrics})
\]

This procedure then creates a series of levels out of the two convenience labels that form the factor PartAuxMetrics. In this way, the string serialization is always in the factus est variant, i.e the order is “normalized”. The first six levels are given below as an example.

42Adams (2013) is also dubious about the strength of the effect of clausula on word order permutations in AVC of the infinitive + auxiliary type.

43It is pointless to appeal to “the right sonority”, as Dyck (2003) puts it, because we have no idea how to measure what “the right sonority” actually is. It is subjective and therefore unoperationalizable.
9.5. CLAUSULAE

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Level</th>
<th>Example scansion</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLLH</td>
<td>pyrrhus-iamb</td>
<td>dátus ērat</td>
<td>57</td>
</tr>
<tr>
<td>HLLH</td>
<td>trochee-iamb</td>
<td>fácēs ērat</td>
<td>129</td>
</tr>
<tr>
<td>LHHH</td>
<td>double-iamb</td>
<td>posīti sūmūs</td>
<td>214</td>
</tr>
<tr>
<td>HHHH</td>
<td>double-spondee</td>
<td>intellect(um) ēst</td>
<td>451</td>
</tr>
<tr>
<td>LLIH</td>
<td>pyrrhus-spondee</td>
<td>sītūs ēssēt</td>
<td>58</td>
</tr>
<tr>
<td>HLIH</td>
<td>spondee-iamb</td>
<td>adfectī sūmūs</td>
<td>346</td>
</tr>
<tr>
<td>LIHH</td>
<td>iamb-spondee</td>
<td>aŭrōgāt(um) ēssēt</td>
<td>239</td>
</tr>
<tr>
<td>HLHH</td>
<td>trochee-spondee</td>
<td>profectīs ēssēt</td>
<td>217</td>
</tr>
<tr>
<td>≤ 3 syllables</td>
<td>NA</td>
<td>fact(um) ēst</td>
<td>698</td>
</tr>
</tbody>
</table>

Table 9.40: Levels for the clausulae predictor PartAuxMetrics

```
levels(pac$PartAuxMetrics)[1:6]
1       abductae.sint
2       ablata.sunt
3       ablatae.sunt
4       ablatus.esset
5       abrogatum.esset
6       absoluta.esse
```

Each level \( k = 1617 \) within the variable was then scanned by hand for its metrical structure, starting with the first syllable in the \( \text{pac} \) sequence and proceeding to the final syllable in the \( \text{pac} \) sequence. Syllables in elidable sequences were treated as such.\(^{44}\) Thus, a level such as \textit{profectus.ērat} scans as \{LHLLH\} (all final syllables in each sequence are marked as being heavy, through convention). R scripts were written to bring down the amount of manual annotation. Through this procedure, we obtain 78 full metrical structures.

Recurring metrical sequences were then identified with the help of material in Aili (1979: 136, Table A1). Aili divides up 32 metrical sequences into eight groups according to the metrical quantity of a sequence’s last four syllables. Specifically, we can identify: (1) pyrrhus-iamb (LL|LH); (2) trochee-iamb (HL|LH); (3) double iamb (LH|LH); (4) spondee-iamb (HH|LH); (5) pyrrhus-spondee (LL|HH); (6) trochee-spondee, a.k.a. double trochee (HL|HH); (7) iamb-spondee (HH|HH); and (8) double spondee (HH|HH). The levels \( k = 78 \) within the variable were then rewritten with this metrical sequence information. Sequences with three syllables or fewer were considered to be unclausulaic (or at least they do not have, or we have not scanned, enough syllables to determine their clausulaic structure), and were accordingly rewritten as missing observations:

```
levels(pac$PartAuxMetrics)[nchar(levels(pac$PartAuxMetrics))<=3]<-NA
```

In sum, there are \( n = 698 \) missing observations. Table 9.40 gives examples of the levels for the variable, along with summary statistics.

9.5.4 Statistical modelling

The overall distribution of this variable and serialization choice is reported in Table 9.41, and Table 9.41 is a complementary barplot. A \( \chi^2 \)-test of independence shows that the distribution is certainly not random \( (\chi^2 = 65.8992, \ df = 7, \ p = 9.917e-12) \). For example, we can see that when the normalized \( \text{pac} \) order scans as a trochee-spondee (i.e. double-trochee), the rate of it being actually realized in that order is quite high (at a rate of 73.27%), but by contrast when the normalized \( \text{pac} \) order scans as a trochee-iamb, the rate of it being realized in that order is quite low (at a rate of 35.66%).

\(^{44}\)It is worth pointing out to facilitate coding, I did not take into consideration the last preverbal word or the following context. It is plausible that left- and right-edge effects may be important, and I accept that this may be a flaw with the present operationalization, but I leave this for a more thorough study of the interaction of clausulae and word order.
Despite the significance of the $\chi^2$-test we need to model this predictor appropriately. However, before we can do so, a variance components model specific to this variable needs to be constructed, because there are 698 missing observations. The first model to be constructed is the basic single-level \texttt{glm}. The second model is a variance components model with varying intercepts for lemma. The third model is a variance components model with varying intercepts for text. The fourth model is a variance components model with varying intercepts for both lemma and text. Table 9.42 shows the deviance and \textit{AIC} values for the four models considered. All three \texttt{vcm}s are significantly better than a null \texttt{glm} which only has the base intercept. Additionally, the two nested \texttt{vcm}s are significantly poorer in terms of the deviance compared with the more complex \texttt{vcm} that contains varying intercepts for both lemma and text. We therefore accept the last model in Table 9.42 as the baseline \texttt{vcm} for this particular predictor.

<table>
<thead>
<tr>
<th>Model</th>
<th>$-2LL$</th>
<th>$AIC$</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{glm} null</td>
<td>2350.133</td>
<td>2352.133</td>
</tr>
<tr>
<td>\texttt{vcm} + varying lemma intercepts</td>
<td>2334.478</td>
<td>2338.478</td>
</tr>
<tr>
<td>\texttt{vcm} + varying text intercepts</td>
<td>2274.211</td>
<td>2278.211</td>
</tr>
<tr>
<td>\texttt{vcm} + varying lemma and text intercepts</td>
<td>2259.671</td>
<td>2265.671</td>
</tr>
</tbody>
</table>

Table 9.42: Constructing a baseline \texttt{vcm} for clausula effects

With the baseline model now in place, let us focus on deriving a statistically adequate model that includes the predictor. Ideally, four models would be constructed and compared. First, a varying
intercepts model would be constructed with the clausula predictor:

```r
clausula.ri <- glmer(Serialization ~ PartAuxMetrics + (1 | Lemma) + (1 | Text), data = clause, family = binomial)
```

In the second model, we would include both varying intercepts and varying slopes for lemma and text:

```r
clausula.rs.both <- glmer(Serialization ~ PartAuxMetrics + (PartAuxMetrics | Lemma) + (PartAuxMetrics | Text), data = clause, family = binomial)
```

The third model would contain varying intercepts for lemma and text, and varying slopes for only lemma:

```r
clausula.rs.lemma <- glmer(Serialization ~ PartAuxMetrics + (PartAuxMetrics | Lemma) + (1 | Text), data = clause, family = binomial)
```

The fourth and final model would contain varying intercepts for lemma and text, and varying slopes for only text:

```r
clausula.rs.text <- glmer(Serialization ~ PartAuxMetrics + (1 | Lemma) + (PartAuxMetrics | Text), data = clause, family = binomial)
```

However, the three more complex models failed to converge after 10000 evaluations. It is possibly
to toy around with the evaluation threshold, for instance increasing the optimizer to 30000 evaluations.
However, even increasing the number did not help in this instance, and in any case it is unwise as it
can lead to very dubious parameter estimates. We will consequently revert to the more simple multilevel
model `clausula.ri`, which converged quite happily, and compare that with the baseline `vcm`. The ANOVA
statistics are reported in Table 9.43. The difference in deviance $63.19 = 2259.67 - 2196.48$ is significantly
closer to zero on $7 = 10 - 3$ degrees of freedom ($p = 3.47169e-11$). We can therefore conclude with
confidence that adding information about clausula vastly increases our explanation of PAC serialization
choice.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>clause.vcm.both</td>
<td>3</td>
<td>2265.67</td>
<td>2282.01</td>
<td>-1129.84</td>
<td>2259.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clausula.ri</td>
<td>10</td>
<td>2216.48</td>
<td>2270.93</td>
<td>-1098.24</td>
<td>2196.48</td>
<td>63.19</td>
<td>7</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 9.43: ANOVA of baseline VCM and GLMER containing clausula information

Let us now walk through the parameter estimates for this clausalae predictor, reported in Table 9.44.
The response variable has been coded 0 for `est factus` and 1 for `factus est`; the logit probabilities in the
estimates are therefore in terms of successes for `factus est` serializations. The predictor has also been
coded using (0-1)-coding, with the level `double-spondee` coded as the baseline level because it contains
the largest number of observations, resulting in greater statistical power. All other levels are thereby
compared to it.

|                     | Estimate | Std. Error | z value | Pr(>|z|) |
|---------------------|----------|------------|---------|----------|
| (Intercept)         | -0.3447  | 0.1817     | -1.896  | 0.05791  |
| PartAuxMetricsspondee-iamb | 0.5149  | 0.1591     | 3.237   | 0.00121  |
| PartAuxMetricstriochee-spondee | 1.2597 | 0.1952     | 6.453   | 1.09e-10 |
| PartAuxMetricsiamb-spondee | 0.2943  | 0.1932     | 1.561   | 0.11864  |
| PartAuxMetricspyrrhus-spondee | -0.3546 | 0.2272     | -1.561  | 0.11864  |
| PartAuxMetricspyrrhus-iamb | 0.7536  | 0.3283     | 2.295   | 0.02171  |
| PartAuxMetricstriochee-iamb | 0.8333  | 0.3265     | 2.552   | 0.01070  |

Table 9.44: GLMER of serialization ~ clausula in normalized PAC order

There are non-significant coefficients. Specifically, iamb-spondee, double-iamb, and trochee-iamb
sequences do not differ significantly from double-spondee sequences. However, there are four significant
coefficients. When the normalized PAC sequence is a spondee-iamb (HHLH) as opposed to a double-spondee sequence, the logit probability of a factus est outcome increases by 0.5149, equating to an odds ratio of 1.673471. When the normalized PAC sequence is a trochee-spondee (HLHH, a.k.a. double trochee), the logit probability of a factus est outcome increases by 1.2597, equating to an odds ratio of 3.524364. When the normalized PAC sequence is a pyrrhus-iamb, the logit probability of a factus est outcome increases by 0.7536, equating to an odds ratio of 2.124635. Finally, when the normalized PAC sequence is a pyrrhus-spondee, the logit probability of a factus est outcome increases by 0.8333, equating to an odds ratio of 2.300899. The confidence intervals (in logits and odds) are given in Table 9.45 for the reader’s convenience.

<table>
<thead>
<tr>
<th></th>
<th>2.5 %</th>
<th>97.5 %</th>
<th>2.5 %</th>
<th>97.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.70</td>
<td>0.01</td>
<td>0.50</td>
<td>1.01</td>
</tr>
<tr>
<td>PartAuxMetricsspondee-iamb</td>
<td>0.20</td>
<td>0.83</td>
<td>1.23</td>
<td>2.29</td>
</tr>
<tr>
<td>PartAuxMetricsstrochee-spondee</td>
<td>0.88</td>
<td>1.64</td>
<td>2.40</td>
<td>5.17</td>
</tr>
<tr>
<td>PartAuxMetricsiamb-spondee</td>
<td>-0.06</td>
<td>0.65</td>
<td>0.94</td>
<td>1.92</td>
</tr>
<tr>
<td>PartAuxMetricsdouble-iamb</td>
<td>-0.13</td>
<td>0.63</td>
<td>0.88</td>
<td>1.87</td>
</tr>
<tr>
<td>PartAuxMetricsstrochee-iamb</td>
<td>-0.80</td>
<td>0.09</td>
<td>0.45</td>
<td>1.10</td>
</tr>
<tr>
<td>PartAuxMetricspyrrhus-iamb</td>
<td>0.11</td>
<td>1.40</td>
<td>1.12</td>
<td>4.04</td>
</tr>
<tr>
<td>PartAuxMetricspyrrhus-spondee</td>
<td>0.19</td>
<td>1.47</td>
<td>1.21</td>
<td>4.36</td>
</tr>
</tbody>
</table>

Table 9.45: Confidence intervals for the model glmer of serialization ~ clausula in normalized PAC order

9.5.5 Discussion

Let us now turn to a more qualitative discussion of the statistical findings presented in the previous section. In qualitative terms, we can say that the metrical sequences to the right of the hierarchy favour the factus est realization, whereas those on the left favour the est factus serialization.

(39) trochee-iamb < double-spondee < double-iamb < iamb-spondee < spondee-iamb < pyrrhus-iamb < pyrrhus-spondee < trochee-spondee

To give an example, if a PAC scans in its normalized order as a trochee-spondee, then it is highly likely to be realized as factus est. By contrast, why should a PAC scanned as a double-spondee or as a trochee-iamb sequence in its normalized order disprefer a factus est serializations? Very basically, one might say that, for example, double-spondees are avoided in prose because they evoke epic poetry, hence the word order switch to est factus; on the other hand, trochee-spondees were a favoured type of clausula, so if the normalized PAC scans as such, an author would routinely actually serialize it in that order, i.e. factus est. This was all noted above, and the evidence here seems to support it. However, this is not an explanation, it is simply a description. I see two linguistic reasons for this distinction.

First, metrical optimization, akin to what we have seen for rhythmical (accentual) and segmental optimization, may be at play. For instance, double-spondee sequences are quantitatively suboptimal: they result in a long span of heavy syllables. By contrast, trochee-spondee sequences are perfectly quantitatively harmonious, in that there is an alternating sequence of light and heavy syllables (ignoring the metrical quantity of the final syllable, which doesn’t count anyway). Writers use word order to optimize metrical structure, that is a word order will be used if it is of optimal metricality and will be dispreferred if it is of suboptimal metricality.

Another interpretation is that certain metrical sequences were cognitively and psycholinguistically easier to access and produce. Rapid metrical parsing of a factus est sequence would with a cognitively easy-to-access-and-produce metrical sequence allow it to be produced more quickly. By contrast, if there is difficult in parsing it (for instance, the participle), then the parsing of a relatively easy structure, i.e. the auxiliary, will help to buy time.
9.6. FINAL REMARKS

In addition to these findings it is worth dwelling on scholars’ use of clausulae to help decide on word order variants when there are textual difficulties in the transmitted texts. While this method gains some support from the current analysis, it is to be noticed that clausulae effects account for a very small amount of the variance in the response variable. Variance explained is typically quantified by Nagelkerke’s $R^2$ for binomially distributed response variables. There is no effective $R^2$ measure for multilevel models, so we will revert to a standard GLM here. That standard GLM shows that clausulae effects account for a mere 5.2% of the variance, even when other things are not controlled for. Consequently, I submit that editors simply cannot rely on clausula alone as the be all and end all of deciding between textual variants. My own views on how such an analysis like the present one can help with textual decisions will be addressed in Part V.

Finally, we briefly mentioned in Chapter 5 that metrical constraints are of critical importance in poetry. They are clearly of much less influence in prose.

9.6 Final remarks

In this chapter we have adduced a great deal of bivariate evidence to demonstrate that phonological factors bear decisively on PAC serialization choice. A summary of the values/levels of each predictor that are associated with each order is given in Table 9.46.

As noted in Section 9.1, debates currently rage about the architecture of grammar, particularly as concerns the position of phonology within it. Some have argued that syntax drives phonology, but not the other way around, and this is the unidirectional, modular view of grammar (“Phonology-Free Syntax”). Others have argued that it can be influential, and that grammar is not modular, with phonology feeding the syntax also. I hope to have shown in this chapter that there is at least some support that phonology impinges on syntax, but whether this can be explained away by theories of modular architectures, I leave for others to debate.

Additionally I wish to stress that, while there are some effects of phonology, many behaving in line with theoretical expections, they make up only a small subset of the information space; furthermore, the effects we have seen are only slight. One cannot build a sufficient description of the PAC alternation based on a group of variables that concern only a single linguistic subdiscipline, as it still leaves a lot of the variation to be explained. We move on in the next chapter to discuss predictors that concern another dimension of formal structure — namely grammar.
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Values/Levels</th>
<th>Outcome</th>
<th>Predictor</th>
<th>Values/Levels</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-clitic</td>
<td>factus est</td>
<td>Clisis and prosody</td>
<td>Auxiliary/Participle Boundary</td>
<td>differing &gt; identical</td>
<td>Rhythmical Optimization</td>
</tr>
<tr>
<td>position 1P</td>
<td>other &gt; 2P</td>
<td>Segmentation</td>
<td>participant/participle Boundary</td>
<td>similar &gt; different</td>
<td>Preverbal/Participle Boundary</td>
</tr>
<tr>
<td>position non-clitic</td>
<td>identical &gt; elidable</td>
<td>Rhythmical Optimization</td>
<td>participant/participle Boundary</td>
<td>similar &gt; different</td>
<td>Preverbal/Participle Boundary</td>
</tr>
<tr>
<td>position</td>
<td>non-clitic</td>
<td>Class and prosody</td>
<td>participant/participle Boundary</td>
<td>similar &gt; different</td>
<td>Preverbal/Participle Boundary</td>
</tr>
</tbody>
</table>

Table 9.46: Summary of phonological and prosodic predictors.
Chapter 10

Grammatical Predictors

10.1 Introduction

In this chapter I discuss grammatical predictors, concerning morphology, morphosemantic/morphosyntactic contrasts, and syntax. The structure of this chapter is as follows. We begin by focusing attention on core aspects of verbal morphology that characterize the construction (Sections 10.2–10.7). Section 10.2 examines the extent to which verbal morphological properties of the auxiliary bear on serialization choice. Two sister sections then follow, with Section 10.3 likewise looking at the morphological TAM structure of the participle, and Section 10.4 examining how prepositional prefixation on the participle might play a role. The next two sections concern the morphosyntactic features of impersonality (Section 10.5) and deponency (Section 10.6). The final section concerning verbal morphology examines whether the conjugational class to which the verb belongs has an effect. After this, we explore how features of nominal morphology that are encoded on the PAC complex, such as number (Section 10.8), gender (Section 10.9), and person (Section 10.10), impinge upon serialization choice. The final two sections — Section 10.11 and Section 10.12 — discuss constituency/grammatical class and clause type, respectively.

10.2 Verbal morphology of the auxiliary

10.2.1 Introduction

Perhaps the most intuitively obvious source of morphosyntactic variation concerns the verbal morphology of the auxiliary, namely its tense, aspect, mood, and finiteness properties.¹

The first dimension of verbal morphology is the tense/aspect of the auxiliary. From the traditional Latin grammars, we can identify several “tenses” of the auxiliary that combine with participles: present tense in (1a), imperfect tense/aspect in (1b), perfect tense/aspect in (1c), pluperfect tense/aspect in (1d), future tense in (1e), and future perfect tense/aspect in (1f).

(1) a. profectus est
    depart.PTCP.REF.PRG.SG.NOM be.3SG.PRS.IND

b. profectus erat
    depart.PTCP.REF.PRG.SG.NOM be.3SG.IMPF.IND

c. profectus fuit
    depart.PTCP.REF.PRG.SG.NOM be.3SG.PF.IND

¹ Nominal morphological features which permeate to the auxiliary and affect its inflectional status are dealt with separately, below.
d. profectus fuerat
depart.PTCPL.PRF.M.SG.NOM be.3SG.PLPRF.IND
e. profectus erit
depart.PTCPL.PRF.M.SG.NOM be.3SG.FUT.IND
f. profectus fuerit
depart.PTCPL.PRF.M.SG.NOM be.3SG.FUTPRF.IND

It is important to note that the tense/aspectual nature of the auxiliary can have real semantic value in certain combinations with the participle: for instance, with future participles, gerundives, adjectives, and adjectival participles the tense has real semantic value. However, with perfect participles, such as the strongly eventive *profectus* ‘departed’ in (1), the auxiliary does not necessarily have real semantic value, and instead the semantics are computed off the entire PAC context.

A second dimension of verbal morphology concerns the mood of the auxiliary: in the examples in (1) the auxiliary is inflected in the indicative, whilst in (2) the auxiliary has a subjective form.

(2) profectus esset
depart.PTCPL.PRF.M.SG.NOM be.3SG.IMP.SBJV

We can also distinguish morphological variability concerning the finiteness of the auxiliary: in the examples in (1)–(2) the auxiliary is finite, whilst in (3) it is nonfinite, specifically an infinitive form.

(3) profectus esse
depart.PTCPL.PRF.M.SG.NOM be.PRS.INF

It will therefore be of interest to examine whether such grammatical properties play a role in the alternation.

10.2.2 Grammatical effects

10.2.2.1 General effects

According to Siewierska (1988: 95–97), word order variation that is influenced by differences in tense/aspect, for instance, is “quite rare”, however she notes exceptions in some African languages and in Chamorro (Austronesian, Malayo-Polynesian). Lendu (Nilo-Saharan, Central Sudanic), for instance, has svo order in perfectives, e.g. *Má a oú ‘I eat chicken’* (Siewierska 1988, 95, ex. 2.150a), and sov order in imperfectives, e.g. *Má oú ‘I am eating chicken’* (Siewierska 1988, 96, ex. 2.150b).

Additionally, effects of finiteness have been reported. For instance, Pollock (1989) observed that there are word order differences between finite and nonfinite forms of the verb in French. It is demonstrated that finite forms raise higher with respect to negation than non-finite forms of the verb. In (4a), the finite verb *aime* raises higher than the negation particle *pas*, whereas in (4b) it does not.2

(4) a. Jean n’aime pas Marie
   John like.FIN NEG Mary
   b. Ne pas aimer Marie
      NEG like.NFIN Mary

The above effects are essentially categorical in nature. However, the statistical gradience of the effect of different verbal inflectional forms can be discerned from a paper by Lohmann (2011) on English infinitive marker omission. He shows that the probability of (e.g.) *help to do* vs *help do* is highest when

---

2 At least this is one way to look at it. One could alternatively say that the negative is displaced, but that does not change the point that finiteness is important to the word order differences in the examples in (4).
the inflectional form of the higher verb is in its uninflected form (i.e. help). By contrast, when the form of the higher verb is a gerund (i.e. helping) as opposed to the uninflected form, the probability of an overtly marked infinitive decreases strongly (odds ratio = 0.29) and significantly so. Similarly, when the form of the higher verb is a past participle or past tense form (i.e. helped) as opposed to the uninflected form, the probability of an overtly marked infinitive also decreases (odds ratio = 0.73), but here the effect is only marginally significant.\(^3\)

10.2.2.2 \(\text{AVC/PAC specific effects}\)

Within works that deal with AVCS outside Latin, tense and other verbal inflectional features are determined not to be an important factor: “a preliminary study showed that [verb tense] did not have a significant impact on the choice of word order” (De Sutter, 2009: 234). By contrast, for Latin PACs, tense has been argued to be of influence.

In his study of the position of verbs in Apuleius (born 125 AD), Möbitz (1923: 24–25) identifies an effect of the auxiliary’s grammatical tense form on the choice of serialization:

\[
\begin{align*}
\end{align*}
\]

The same effect is also claimed to be in operation in the works of St. Augustine (Muldowney, 1937: 133f.), even though he is a much later writer (354–430 AD). Muldowney (1937) states: “an examination of the instances of [est factus] shows several instances with the perfect tense of the copula”.

In view of these claims, we would be more likely to find (5a) instead of (5b) and (6a) instead of (6b) in these texts.

(5) a. Ingens **exin oborta est** epulonum
   huge.F.SG.NOM then arise.PTCPL.PFR.F.SG.NOM be.3SG.PRS.IND banqueter.M.PL.GEN
   expectatio suspende.F.SG.NOM

   b. Ingens **exin est oborta** epulonum
   huge.F.SG.NOM then be.3SG.PRS.IND arise.PTCPL.PRF.F.SG.NOM banqueter.M.PL.GEN
   expectatio suspende.F.SG.NOM
   ‘Then the huge suspense of the banqueters arose’ (cf. Apul. Met. 10. 16)

(6) a. uti **fuerat pollicitus**
   just.as be.3SG.PL.PRF.IND promise.PTCPL.PFR.M.SG.NOM

   b. uti **pollicitus fuerat**
   just.as promise.PTCPL.PFR.M.SG.NOM be.3SG.PL.PRF.IND
   ‘Just as he had promised’ (cf. Apul. Met. 7. 5)

In sum to this section, there are cross-constructional and cross-linguistic reasons for including a predictor for the (verbal) inflectional morphology of the auxiliary. However, as yet, nothing has been quantified with respect to this variable for the PAC alternation.

\(^3\)The inflectional form helps also behaves differently from help, but this is a distinction in grammatical number and so irrelevant to this present discussion.
10.2.3 Variable profile and operationalization

Against the above background a predictor variable is included for the PAC’s verbal morphology (MorphologyAuxiliary). The coding was automated by running an R script over the auxiliary’s word form. To reduce data sparsity it was decided to collapse levels with fewer than \( n < 20 \) observations recorded on them. Specifically, for present tenses a distinction was made between present indicatives (presentIndicative, \( n = 1279 \)), present subjunctives (presentSubjunctive, \( n = 197 \)), and present infinitives (presentInfinitive, \( n = 248 \)). For imperfects, a distinction was made only between indicatives (imperfectIndicative, \( n = 274 \)) and subjunctives (imperfectSubjunctive, \( n = 314 \)), the only formal distinction that can be made anyway. For perfects, only tense/aspect information was used (perfect, \( n = 48 \)), as was done for pluperfects (pluperfect, \( n = 20 \)). Futures, future perfects, and future infinitives were collapsed into a single level (future, \( n = 29 \)).

10.2.4 Statistical modelling

The crossclassification of this variable by serialization choice is given in Table 10.1. This suggests some interesting correlations. First, present indicatives show the highest rate of factus est serializations (71.07%), whereas imperfect indicatives and pluperfects show the lowest rates (31.39% and 35%, respectively). In between these extremes, present subjunctives and present infinitives, for example, appear to favour factus est only slightly. However, a statistical modelling approach is clearly called for.

<table>
<thead>
<tr>
<th></th>
<th>Serialization est factus (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>present indicative</td>
<td>370 (28.93)</td>
<td>909 (71.07)</td>
<td>1279</td>
</tr>
<tr>
<td>imperfect indicative</td>
<td>188 (68.61)</td>
<td>86 (31.39)</td>
<td>274</td>
</tr>
<tr>
<td>imperfect subjunctive</td>
<td>137 (43.63)</td>
<td>177 (56.37)</td>
<td>314</td>
</tr>
<tr>
<td>perfect</td>
<td>18 (37.50)</td>
<td>30 (62.50)</td>
<td>48</td>
</tr>
<tr>
<td>pluperfect</td>
<td>13 (65.00)</td>
<td>7 (35.00)</td>
<td>20</td>
</tr>
<tr>
<td>present infinitive</td>
<td>119 (47.98)</td>
<td>129 (52.02)</td>
<td>248</td>
</tr>
<tr>
<td>present subjunctive</td>
<td>97 (49.24)</td>
<td>(100)</td>
<td>197</td>
</tr>
<tr>
<td>future</td>
<td>17 (58.62)</td>
<td>12 (41.38)</td>
<td>29</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
<td>2409</td>
</tr>
</tbody>
</table>

Table 10.1: Relationship between the verbal inflection of the auxiliary and PAC serialization

To model this variability, a single multilevel model is considered which includes the predictor term and varying intercepts for lemma and text:

```
morphaux.ri<-glmer(RealizationOfPAC~MorphologyAuxiliary+(1|LemmaParticiple)+(1|Text),pac,family=binomial)
```

No other multilevel models are considered, because of model evaluation issues when random slopes are included. This model is then compared against the baseline vcm via an ANOVA, as presented in Table 10.2. This demonstrates that the inclusion of the predictor is required: the deviance is significantly reduced from 3139.41 to 2992.98, a statistically significant difference of 160.43 on a \( \chi^2 \)-distribution with \( 10 - 3 = 7 \) degrees of freedom.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>ChiDf</th>
<th>Pr(&gt;Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.vcm</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>morphaux.ri</td>
<td>10</td>
<td>2992.98</td>
<td>3050.85</td>
<td>-1486.49</td>
<td>2972.98</td>
<td>160.43</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 10.2: ANOVA of baseline glm and selected glmer containing grammatical number information on the subject NP

It is therefore important to inspect the parameters of this model to examine the direction of the effects (Table 10.3). In this model, factus est is coded as a successful response (=1), and the predictor variable
is coded such that present indicatives (the largest category) constitute the baseline level to which the rest are to be compared. The intercept therefore indicates the logit probability of a factus est outcome when the auxiliary is a present indicative, which is 0.71 equivalent to 2.04 in odds or a probability of 0.67 (cf. the similarity between this probability and the percentage in Table 10.1; the differences are because baseline lemma and textual fluctuations have been controlled for). All the coefficients are statistically significant, barring the level for perfects. Let us discuss the statistically significant effects in terms of their increasing effect sizes (which are all in negative logit space): imperfect subjunctives, present subjunctives, present infinitives, futures, pluperfects, and imperfect indicatives. First, when the auxiliary is an imperfect subjunctive compared to when it is a present indicative, the logit probability of a factus est outcome decreases by $-0.5445$ (an odds ratio of 0.580146), and the confidence intervals associated with this estimate lie between 0.44334011 and 0.7591639 on the logit scale. Second, when the auxiliary is a present subjunctive compared to when it is a present indicative, the logit probability of a factus est outcome decreases by $-0.8430$ (an odds ratio of 0.374045), and the confidence intervals associated with this estimate lie between 0.30937796 and 0.5988310 on the logit scale. Third, when the auxiliary is a present infinitive compared to when it is a present indicative, the logit probability of a factus est outcome decreases by $-0.9834$ (an odds ratio of 0.2576646), and the confidence intervals associated with this estimate lie between 0.27795851 and 0.5033472 on the logit scale. Fourth, when the auxiliary is of a future form compared to when it is a present indicative, the logit probability of a factus est outcome decreases by $-1.3561$ (an odds ratio of 0.2576646), and the confidence intervals associated with this estimate lie between 0.11755183 and 0.5647811 on the logit scale. Fifth, when the auxiliary is of pluperfect form compared to when it is a present indicative, the logit probability of a factus est outcome decreases by $-1.4084$ (an odds ratio of 0.2445337), and the confidence intervals associated with this estimate lie between 0.09313259 and 0.6420603 on the logit scale. Sixth, when the auxiliary is an imperfect indicative compared to when it is a present indicative, the logit probability of a factus est outcome decreases by $-1.6748$ (an odds ratio of 0.1873430), and the confidence intervals associated with this estimate lie between 0.13824572 and 0.2538768 on the logit scale. Notice that for this particular level, the confidence intervals are quite narrow, indicating an extremely reliable result.

| Estimate     | Std. Error | z value | Pr(>|z|) |
|--------------|------------|---------|----------|
| (Intercept)  | 0.7130     | 0.1619  | 4.404    | 1.06e-05 |
| MorphologyAuxiliaryimperfectIndicative | -1.6748 | 0.1551 | -10.801 | 2e-16 |
| MorphologyAuxiliaryimperfectSubjunctive | -0.5445 | 0.1372 | -3.968 | 7.25e-05 |
| MorphologyAuxiliaryperfect | -0.4668 | 0.3222 | -1.449 | 0.14735 |
| MorphologyAuxiliarypluperfect | -1.4084 | 0.4925 | -2.860 | 0.004242 |
| MorphologyAuxiliarypresentInfinitive | -0.9834 | 0.1515 | -6.492 | 8.49e-11 |
| MorphologyAuxiliarypresentSubjunctive | -0.8430 | 0.1685 | -5.004 | 5.63e-07 |
| MorphologyAuxiliaryfuture | -1.3561 | 0.4004 | -3.387 | 0.00707 |

Table 10.3: Model parameters of final glmer for serialization ~ auxiliary’s verbal morphology

10.2.5 Discussion

To turn to a discussion, it appears that the auxiliary’s verbal morphology does have an effect on the alternation, which had been claimed before in the literature. However, the effects do not entirely tie in with previous claims. First, perfect auxiliaries (i.e those with a fu- root) only behave as claimed when the auxiliary is specifically a pluperfect form (e.g. fuerat), but crucially not when it is a perfect form (e.g. fuit). It is possible that different effects might be found when one examines later texts.

The results suggest several new findings. First, it is important that imperfect forms, both indicative and subjunctive, future forms, and present subjunctives and indicatives behave strikingly differently to present indicatives. Second, using the information of the logit probability estimates, a cline of probabilistic tendencies can be established regarding verbal inflectional forms most likely to be found with factus est.
(cf. (7)).

(7) present indicative > perfect > imperfect subjunctive > present subjunctive > present infinitive > future > pluperfect > imperfect indicative

Why is the auxiliary’s verbal morphology important in the way characterized in (7)? One possibility is that because the different inflectional forms of the auxiliary have different metrical structures, the factor of clausula — as mentioned earlier — may be a determining factor. Indeed, Möbitz (1923) mentions that phonological factors may be in play. He states (1923: 125) that the sequence of a fu-auxiliary followed by a participle is particularly effective in ensuring a rhythmical closure for the clause (i.e. clausula).

Another possibility concerning phonology is that clisis is overriding in influential: for instance, the majority of present indicatives (68.73%) are of the form est, which was identified above as being a phonological phrase clitic and strongly preferring factus est. By contrast, imperfect indicatives may be more robustly intonational phrase clitics, preferring est factus when the PAC is docked in the appropriate phrasal position; alternatively, they might not be clitics of any type, being free to vary between est factus and factus est orders.

It may also be that frequency or complexity factor into the mix. However, while it is for the most part straightforward to establish links between complexity and the levels of binary variables, it is much less clear how to go about doing so for factors with many levels, such as the present one. One could say that present indicative forms such as sunt are easier to access than imperfect indicative such as erant, but that can only remain a speculation. However, the present research does suggest that if complexity is important, then morphological differences in living languages might be worth investigation to at least substantiate this speculation with a bit more clarity.

Finally, the verbal inflection of the auxiliary is not independent of clause type: auxiliaries are indicative in main clause declaratives, but they can be subjunctive in (e.g.) cum-clauses. We will explore the relevance of clause type below.
10.3 Morphosyntactic category of the participle

10.3.1 Introduction

Similarly to the previous variable, another source of morphological variation in the present dataset concerns the morphological category of the participle. The participles sampled from include perfect participles (either as verbal participles, adjectival participles, or as pure adjectives), future participles, and gerundives. Might this type of variation play an important role in PAC serialization choices?

10.3.2 Grammatical effects in AVCs

There has been some discussion in the AVC literature that the morphological properties of the nonfinite predicate (e.g. participle, infinitive) can influence word order variability. However, there are mixed findings for Latin, which may be a result of data sample choices.

10.3.2.1 Dutch

In Dutch verb cluster constructions involving the verb *zijn* ‘to be’, a distinction is made between constructions such as *dat de winkel gesloten is* ‘that the shop is closed’, where the participle *gesloten* is an adjectival participle with the distribution of an adjective and the finite verb is receives a copula interpretation, and *dat de winkel gesloten is* ‘that the shop has been closed’, where the participle *gesloten* is a verbal participle and the finite verb *is* receives an auxiliary interpretation: the former “[focuses] on the durative state of being closed . . . while the latter [focuses] on the action of closing the shop” (De Sutter 2009, 236, incl. examples). This distinction has been claimed, and is found to have, an effect on the order of the participle and *zijn*. De Sutter (2005, 2009) cites grammatical and semantic-pragmatic literature where it claimed that adjectival participles appear exclusively in the participle-*zijn* sequence, whereas verbal participles can appear either in the *zijn*-participle sequence or in the participle-*zijn* sequence. Basing himself on nearly 2400 data-points, De Sutter finds strong effects for this distinction in both his univariate and multivariate analyses. The effects are in the expected direction; however, he shows that adjectival participles do not exclusively favour the participle-*zijn* sequence, as had previously been asserted.

10.3.2.2 Latin

Turning now to Latin PACs, in Brookes (2009) I claimed that the three-way distinction between the different types of participles (i.e. perfect participles vs. future participles vs. gerundives) has an impact on the choice of serialization. Using data from Tacitus, I said that in the *factus est* variant the perfect participle is “distinctive”, while in the *est factus* variant the gerundive is “distinctive” (2009: 60). This “distinction” is apparently because, although the perfect participle is common overall for both variants, the gerundive is relatively more common with the *est factus* variant than it is with the *factus est* variant. Drawing on the data in my appendix there (and highlighting only the relevant serializations here5), in the *factus est* variant around 90% (658) of participles are of the perfect type, about 7% (51) are of the gerundive type, and around 3% (24) are of the future type. When it comes to the *est factus* variant, approximately 86% (12) of participles are of the perfect type, 14% (2) of participles are of the gerundive type, and 0% (0) appear in the future type. Because of the percentage difference, with 7% of PACs being assigned to the gerundive in the *factus est* variant and 14% in the *est factus* variant, I claimed that for Tacitus the gerundive participle was “distinctive” in the latter serialization. But I have since changed my mind, because statistical analyses were not used in that investigation. Conducting a Fisher exact

---

4Nominal morphological features which permeate to the participle are dealt with separately, below.
5As noted earlier, in Brookes (2009) I examined serialization variability with respect to both continuous and discontinuous serializations.
test over the data in the 2009 study, the differences are not significant \((p = 0.5422)\), and so one could not really claim that there is any relationship between participle type and the variant used. In addition to this, authors seem to differ in the choice of “distinctive” participle for the various serializations: in Sallust, Livy, and Cicero, it is the perfect participle that associates with \(est \text{ factus}\), showing the opposite trend.

Muldowney (1937: 129–135), examining word order in St. Augustine, finds that 77.42664% of (again unspecified) participles associate with the \(factus \text{ est}\) order, 80.99174% of gerundives associate with the \(factus \text{ est}\) order \((\chi^2 = 0.5147, df = 1, p = 0.4731, \phi = 0.035)\). However, as Muldowney (1937: 135) notes, the gerundive + \(esse\) construction occurs “with comparative infrequency” in the author she studies; consequently, this may be a reason why strong effects are not found.

In sum, there has been a discussion of the relevance of participial morphology on serialization choice in AVCS across languages. However, the effects differ in their strength depending on the study, the language, and the period in question. Second, there has been little interpretation as to why the morphology of the participle should matter.

10.3.3 Variable profile and operationalization

As discussed in Chapter 3 and 4, the PAC has differing morphologies that overlap and interact, to some extent, with its constructional semantics. We have also seen that some researchers have argued that it might be an influential variable, although little had been explored in quantitative detail. In order to take into account the possibility that this may impinge on the variation, I first distinguish the three broad types of participle considered in this study, based on their formal morphological structure: (i) perfect participle, (ii) future participle, and (iii) gerundive.

Perfect Participle:

\[(8)\]
\[
a. \quad \text{a from quibus \textit{erat} \textit{profectus}} \\
\text{from REL.M.PL.ABL be-3SG.IMPF.IND depart-PTCPL.PRF.M.SG.NOM} \\
\text{‘from whom he had departed’ (Nep. \textit{Milt. 2. 3})}
\]
\[
b. \quad \text{quae \textit{exorsus} \textit{sum}} \\
\text{REL.N.PL.ACC begin-PTCPL.PRF.M.SG.NOM be-1SG.PRS.IND} \\
\text{‘which I have begun’ (Nep. \textit{Praef. 8. 2})}
\]
\[
c. \quad \text{Orgetorix \textit{mortuus} \textit{est}} \\
\text{Orgetorix-M.SG.NOM die-PTCPL.PRF.M.SG.NOM be-3SG.PRS.IND} \\
\text{‘Orgetorix died’ (Caes. \textit{Gal. 1. 4. 3})}
\]

Future Participle:

\[(9)\]
\[
a. \quad \text{ego \textit{istud} \textit{facturus} \textit{sum}} \\
\text{PERS.PRN.1SG.NOM DEM.N.SG.ACC do-PTCPL.FUT.M.SG.NOM be-1SG.PRS.IND} \\
\text{‘I am about to do that’ (\textit{B.Afr. 45. 4})}
\]

Gerundive:

\[(10)\]
\[
\text{si mons \textit{erat} \textit{ascendendus} \ldots} \\
\text{if mountain-M.SG.NOM be-3SG.IMPF.IND climb-GER.M.SG.NOM \ldots} \\
\text{‘If the mountain had to be climbed \ldots’ (Caes. \textit{Civ. 1. 79. 2})}
\]
As noted in Chapter 3 and 4, perfect participles are distinguished via their inflection in the allomorphs -t-, as in (8a) and -s-, as in (8b); as noted there, I include the aberrant form mortuus ‘having died, dead’ as a perfect participle (see (8c)). Likewise, gerundive and future participles are easily identifiable based on their morphology: the former have -nd morphemes, as in the latter -ur- morphemes (of which -tur- and -sur- are the two allomorphs). To facilitate coding I semi-automatically tagged this variable. First, all cells were blindly tagged by default as perfect participle. I then searched the variable column with all the individual word forms of the participle for character strings in nd, tur and sur and replaced the cells where a match was positive with the appropriate participle label, e.g. gerundive or future participle. The search obviously throws up some false positives, such as inductus ‘brought in’, which contains the nd character string, but it is not a gerundive morpheme. Consequently, such cases were disregarded as inappropriate.

However, as noted earlier, perfect participles can differ as to whether they are verbal or adjectival. While identifying distinctions between future participles, gerundives, and perfect participles is straightforward as it is based on the overt morphological structure, it is less easy to identify differences within the perfect participle category, that is whether they are verbal or adjectival. Whilst some cases are intuitively verbal, as in (11a), and some are intuitively adjectival, as in (11b), others are ambiguous, as in (11c):

(11) a. populi scito in patriam restitutus est.
   ‘[Aristides] was restored to his country by a decree of the people’ (Nep. Arist. 1. 5) (verbal)

b. neque eo fuit contentus
   ‘And he was not content with that’ (adjectival) (Nep. Ham. 2. 5)

c. quae in proximo litore erant collocata
   EITHER ‘which had been located on the next shore’ (verbal)
   OR ‘which were located on the next shore’ (adjectival) (Nep. Han. 11. 4)

For example, in (11c), the pac erant collocata can be interpreted as a past perfect construction, in which case the participle is verbal, or as a past resultant state construction, in which case the participle is adjectival. The purpose of this section is to present an annotation effort to assist in categorizing (perfect) participles into one of the above three categories — namely, verbal, adjectival, or ambiguous.

Although the phenomenon of adjectival uses of participles has long been known, it wasn’t really until Wasow’s (1977) important paper which formally identified “two sources for passives” for English (1977: 338) that interest in the phenomenon grew considerably. Since then, a lot of the discussion has revolved around identifying diagnostics that characterize a participle as being verbal or adjectival (cf. Coussé 2011: 613). In addition, researchers working on typologies of resultative constructions envisaged more broadly, of which the be + participle type is just one kind, have identified various contexts in which such constructions prevail (e.g. Nedjalkov & Jaxontov 1988). Third, annotation guides for parsed corpora have also come up with diagnostics to help separate the two (Santorini, 1990). More specialized schemes have attempted to computationalize the categorization (De Sutter, 2005). In what follows, I draw these various related lines of research lines together, and report on the diagnostics and contextual tests have been proposed which allow one to consider a participle, or a construction, an adjectival passive and another a verbal passive. I discuss connections with Latin throughout.

10.3.3.1 Phonomorphology

A number of phonomorphological diagnostics have been mentioned in relation to the verbal vs. adjectival participle distinction. For instance, Dubinsky & Simango (1996: 777-780) show that for English the

6While the former are derived from PIE *-tos, the latter appears to come from the stem *mort- with the participial ending *-wos.
participle marker \(-ed\) in some adjectival participles can be realized with the allophone \([-id]\), while verbal participles are realized as \([-d]\) and \([-t]\), as in blessed \([bl\text{Est}]\) vs. blessed \([bl\text{EssId}]\). That this is the case can be shown by the repulsion of the \(-[d]\) form in prototypical verbal contexts (for the agentive by-phrase, see below), as in (12):

(12) The congregation was blessed \([bl\text{Est}]/*[bl\text{EssId}]\) by the pastor.

(cf. Dubinsky & Simango 1996: 776, ex. 57b, modified)

Another phonomorphological characteristic of English adjectival participles is that some of these can receive the allomorph \(-en\), as in open, shaven, while verbally-used ones have \(-ed\) (cf. Embick 2004: 358 and Sleeman 2011: 1570).

This feature identified for English does not have an exact parallel in Latin, as \(-tus\) and the phonologically conditioned variant \(-sus\) are possible for both adjectival and verbal participles. However, in some instances one finds phonological differences within the participle’s lexical root. For instance, the verbal participle of lauo ‘wash’ is lotus, not lautus, which is reserved for the adjective lemma. The verbal participle of tueor ‘watch over’ is tuitus, not tutus, which is reserved for the adjective. While these differences clearly have an effect on the categorical status of the participle in these two cases, they are not found for all lemma.

10.3.3.2 Morphology

10.3.3.2.1 Negative prefixation A well-known morphological diagnostic is the ability of the adjectival participle to receive privative un-prefixation whereas the verbal participle cannot. For instance, in the bride is loved, the participle loved can be prefixed with un- (OED s.v. unloved). This feature makes the participle a strong candidate for adjectival potential (cf., inter alia, Siegel 1973; Bresnan 1996; Birner et al. 2002; Emonds 2006; Maienborn 2007; Stolterfoht et al. 2010).

This feature is pertinent to the Latin data as well, as a number of participles can receive in-prefixation (OLD s.v. in-\(^2\)), as in for instance imparatus ‘unprepared’ besides paratus ‘prepared’, and inuictus ‘unconquered’ an in-prefixed form of the participle uictus ‘conquered’ from uinco ‘conquer’. However, this is not a cut-and-dried diagnostic as many participles that are listed as adjectives in OLD are not attested with in-prefixation: mortuus is one such example, as there is no participle immortuus ‘undead’.

10.3.3.2.2 Degree modification Degree modification has also been argued to affect only adjectival participles (Wasow, 1977; Levin & Rappaport, 1986; Bresnan, 1996; Birner et al., 2002; Emonds, 2006; Maienborn, 2007; Stolterfoht et al., 2010), because it affects only adjectives (John is very happy) but not verbs (*John very washed the car). For instance, the degree modifier very can modify the adjectival use of John is (very) excited felicitously, but it cannot modify the more verbal types such as John is (*very) departed and John is (*very) killed. The same is true of too, as in John was too excited by the news to do his homework vs. *The window was too broken by the burglar to be able to be fixed (cf. Birner et al. 2012).

This feature is true of Latin adjectival participles as well. Those that are attested with good adjectival background can be found with synthetic comparativization in -ior, as in paratior (← paratus, participle of paro) and synthetic superlativization in -issimus, as in paratissimus. By contrast, we do not find interfectus, the participle of the verb intericio, graded as *interfectior in the comparative and *interfectissimus in the superlative. This points to the verbiness of this participle. However, this diagnostic is not completely straightforward, as a number of participles that are listed as adjectives by OLD cannot be compared or superlativized, e.g. pactus ‘agreed, pledged’.

10.3.3.2.3 Word-formation Stolterfoht et al. (2010: 127,129), writing on German, report that the ability for the participle to be used productively in building compound words is an important diagnostic in
10.3. MORPHOSYNTACTIC CATEGORY OF THE PARTICIPLE

helping to decide whether the participle can be used adjectivally, such as whether the participle can be compounded with an adverb, noun, preposition, or some other complex-word building entity. One example is they give is *Poiger and Cerny sind grippeerkrankt*, where the participle *erkrankt* is morphologically built with the nominal first member *Grippe ‘flu*. This points to the adjectivity of *erkrankt*.

On this approach, a Latin compound formation such as *circumsutus*, built from *sutus* participle of *suo ‘I sew* and the preposition *circum*, would suggest the adjective potentiality of *sutus*. (There is no attested verb *circumsuo*.)

10.3.3.3 Semantics

10.3.3.3.1 Telicity One semantic diagnostic that has been frequently mentioned is the telicity of the predicate (Kratzer, 2001; Emonds, 2006; Coussé, 2011). Telic predicates (i.e. achievements and accomplishments), such as *break* and *open*, include an eventive component, involving a causational phase and/or a processual phase, and a resultant state phase. Because such participles include a resultant state phase, this makes them excellent for use as adjectival participles. By contrast, atelic (activity) predicates such as *carry, drive*, which have no resultant state phase in their denotation, but just a causational phase and/or processual phase, sound awkward as adjectival participles: there is nothing inherent in their meaning that specifies the resultant end point. Thus, the *car is driven* has a verbal interpretation in English. However, things are a bit fuzzier than categorically-minded linguists would like. It has been noticed (by e.g. Kratzer 2001) that activity verbs can be used in resultant state function, when they are given a “job is done”, “activity is over” reading. For example, as Kratzer (2001) notes, cat petting is an inherently atelic predicate, but it can be coerced into a telic predicate under a “job is done” interpretation, for instance, if it was someone’s job to carry out cat petting. Then *the cat is petted* is no longer a verbal passive, but an adjectival construction.

10.3.3.3.2 Specialized meanings Another semantic diagnostic mentioned is the use of a participle with a metaphorical or specialized meaning which is divergent from that of the verb (Birner et al., 2002). If it is so used, then it is unambiguously adjectival. As Birner et al. (2002: 1440) notes for English participles, such as in *She’s bound to win* and *We’re engaged*, “they are no longer comparable to verbal passives with the same forms, and their connection with passives proper is purely historical”.

This feature is applicable to Latin. Some participles have a completely divergent meaning from the related participle of the verb. For instance, *altus ‘having great extension upwards, lofty, tall, high’, contentus ‘content, satisfied’, privatus ‘restricted for the use of a particular person, private’, have meanings distinct from their verbs, *alo ‘to suckle, nurse, feed’, contineo ‘to hold together, connect, link, join’, and privuo ‘to cause to be parted (from), deprive or rob (of)’, respectively.

10.3.3.4 Syntax

10.3.3.4.1 Complementation By far the most important syntactic diagnostic (and probably the most important diagnostic of all) is that of complementation after verbs which can only take adjective phrase complements (e.g. *act, appear, become, look, remain, seem, smell, sound*) (Birner et al., 2002; Emonds, 2006). Consider the use of the participle *broken* in the following examples:

(13) a. The glass looked broken
    b. *The glass looked broken by the burglar*

While (13a) is fine, and is to be interpreted as an adjectival passive, that in (13b) is syntactically ill-formed, and so here *broken* has be interpreted as a verbal passive.

10.3.3.4.2 Coordination I mention the case of coordination (cf. Maienborn 2007; Stolterfoht et al. 2010). If the participle is coordinated with a true adjective, then the participle has to be interpreted adjectivally.

155
Thus, (14a) is fine, but (14b) is odd.

(14) a. Jennifer is happy about her work but worried whether she will remain so.
    b. Jennifer is happy about her work and given a promotion.

Given that in the following Latin example the participle *contractae* is coordinated with an adjective *inexercitatae*, it makes sense to treat the former adjectivally as well, despite the presence of the moment time adverbial.

(15) *quod et inexercitatae et non multo ante erant contractae*

    ‘in that they were untrained and freshly mustered’ (Nep. *Ages*. 3. 3)

10.3.3.4.3 Agentive *by*-phrase As can be gleaned from several of the examples above, the presence of an agentive *by*-phrase renders the construction verbal (Wasow, 1977; Levin & Rappaport, 1986; Birner et al., 2002; Emonds, 2006). Be that as it may, non-agentive *by*-phrases can occur in adjectival passive constructions:

(16) Jennifer seemed worried by the sudden news.

Birner et al. (2002) observe that “by phrases are permitted in adjectival passives when the meaning of the corresponding verb is stative but not when it is dynamic. Adjectival worried by the prospect of redundancy ... is admissible because the corresponding verb, worry, has a stative meaning”. As Birner et al. (2002) note, while *by*-phrases are sometimes not useful, it is often helpful to try and replace the preposition *by* with another preposition. If it can take another preposition and the meaning is roughly the same, it is necessarily adjectival. For instance, in (16), replacing *by* with *about* or *at* is admissible.

10.3.3.5 Summary of diagnostics

In Table 10.4 a summary of the diagnostics discussed in this section are presented. The above diagnostics were used, with the following broad guidelines:

- If the observation passes the adjectivalhood diagnostics but cannot plausibly be used with an active synthetic verbal construction (i.e. there is no verb or it is otherwise impossible or difficult to use an active verb in the context), then it is tagged as being adjectival. For example, in the following examples, it is difficult to assume that an event took place, or else it is presented as being irrelevant, and sounds odd with a ‘they’ subject; it just occurs that way through nature.

  (17) *namque Vlia in edito monte posita est*

    ‘for Ulia is located on a high mountain’ (*B. Alex*. 61. 3)

  (18) *circumpositi sunt huic oppido magni ... colles ‘Many hills surround this town’* (*B. Alex*. 72. 2)

- If the observation fails the adjectival diagnostics, then it is tagged as being necessarily verbal. For instance, if it fails the complementation test (which is a sufficient and necessary condition for adjectivalhood (Birner et al., 2002)), then it has to be interpreted as being verbal, as complementation only allows adjectives.

- If the observation can be used with a PAC that passes adjectivehood diagnostics or with a synthetic verbal construction, then it is necessarily ambiguous. For instance, if it can plausibly be construed with a subject (such as ‘they’ or one derived from the context), then it can be verbal as well as adjectival, and is consequently ambiguous between both readings.
10.3. MORPHOSYNTACTIC CATEGORY OF THE PARTICIPLE

<table>
<thead>
<tr>
<th>Diagnostic set</th>
<th>Diagnostic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonomorphology</td>
<td>Lexical change in root</td>
</tr>
<tr>
<td>Morphology</td>
<td>Negative prefixation</td>
</tr>
<tr>
<td></td>
<td>Degree modification</td>
</tr>
<tr>
<td></td>
<td>Word-formation</td>
</tr>
<tr>
<td>Semantics</td>
<td>Telicity</td>
</tr>
<tr>
<td></td>
<td>Specialized meaning</td>
</tr>
<tr>
<td>Syntax</td>
<td>Complementation</td>
</tr>
<tr>
<td></td>
<td>Coordination</td>
</tr>
<tr>
<td></td>
<td>Agentive by-phrase</td>
</tr>
<tr>
<td></td>
<td>Replacability of by-phrase</td>
</tr>
</tbody>
</table>

Table 10.4: Adjectivehood diagnostics

10.3.4 Statistical modelling

As Table 10.5 suggests, there appears to be an uneven distribution: perfect participles which have the morphosyntactic distribution of verbs exhibit the lowest rate of est factus orders (at 35.20%), whereas perfect participles ambiguous between an adjectival or verbal interpretation show the highest rate of est factus (at 65.64%). This asymmetry is clearly significant on a straightforward \( \chi^2 \)-test of independence (\( \chi^2 = 83.2655, df = 4, p < 0.001 \)).

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>perfect-verbal</td>
<td>610 (35.20)</td>
<td>1123 (64.80)</td>
<td>1733</td>
</tr>
<tr>
<td>future</td>
<td>57 (44.19)</td>
<td>72 (55.81)</td>
<td>129</td>
</tr>
<tr>
<td>gerundive</td>
<td>87 (46.28)</td>
<td>101 (53.72)</td>
<td>188</td>
</tr>
<tr>
<td>perfect-ambiguous</td>
<td>149 (65.64)</td>
<td>78 (34.36)</td>
<td>227</td>
</tr>
<tr>
<td>perfect-adjectival</td>
<td>56 (42.42)</td>
<td>76 (57.58)</td>
<td>132</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
<td>2409</td>
</tr>
</tbody>
</table>

Table 10.5: Relationship between participle morphological category and PAC serialization

To turn to statistical modelling, four multilevel models are built, with the first simply specifying the predictor term and varying intercepts for lemma and text:

```r
partcat.ri<-glmer(RealizationOfPAC~ParticipleCategory+(1|LemmaParticiple)+(1|Text),pac,family=binomial)
```

The second model is the same as the first, but with the addition of varying slopes for lemma and text.

```r
partcat.rs.both<-glmer(RealizationOfPAC~ParticipleCategory+(ParticipleCategory|LemmaParticiple)+(ParticipleCategory|Text),pac,family=binomial)
```

The third model contains varying slopes for only lemma:

```r
partcat.rs.lemma<-glmer(RealizationOfPAC~ParticipleCategory+(ParticipleCategory|LemmaParticiple)+(1|Text),pac,family=binomial)
```

And the fourth model contains varying slopes for only text:

```r
partcat.rs.text<-glmer(RealizationOfPAC~ParticipleCategory+(1|LemmaParticiple)+(ParticipleCategory|Text),pac,family=binomial)
```

The model comparisons in Table 10.6 indicate that `partcat.rs.text` may be both the most informative and parsimonious at the same time, as its AIC is the lowest. This can be confirmed by comparing the −2\( \text{LL} \) of the simplest model with the three complex models. The deviation reduction between that
of \texttt{partcat.ri} and \texttt{partcat.rs.lemma} is not significant ($\chi^2 = 11.437$, $df = 14$, $p = 0.6514$), but that for \texttt{partcat.rs.text} and \texttt{partcat.rs.both} both are ($\chi^2 = 38.744$, $df = 14$, $p = 0.0003995$ and $\chi^2 = 48.071$, $df = 28$, $p = 0.01053$, respectively). Comparing \texttt{partcat.rs.both} with \texttt{partcat.rs.text} and \texttt{partcat.rs.lemma} then demonstrates that adding varying slopes for lemma is superfluous entirely ($\chi^2 = 9.3273$, $df = 14$, $p = 0.8095$), but significant information is lost if varying slopes for text are excluded ($\chi^2 = 36.634$, $df = 14$, $p = 0.000838$). Therefore, it is \texttt{partcat.rs.text} that should be preferred for further appraisal.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>partcat.ri</td>
<td>7</td>
<td>3072.79</td>
<td>3113.30</td>
<td>-1529.40</td>
<td>3058.79</td>
</tr>
<tr>
<td>partcat.rs.lemma</td>
<td>21</td>
<td>3089.35</td>
<td>3210.88</td>
<td>-1523.68</td>
<td>3047.35</td>
</tr>
<tr>
<td>partcat.rs.text</td>
<td>21</td>
<td>3062.05</td>
<td>3183.57</td>
<td>-1510.02</td>
<td>3020.05</td>
</tr>
<tr>
<td>partcat.rs.both</td>
<td>35</td>
<td>3080.72</td>
<td>3283.26</td>
<td>-1505.36</td>
<td>3010.72</td>
</tr>
</tbody>
</table>

Table 10.6: ANOVA statistics for various \texttt{glmer}s containing participle category information

We then compare this model with the baseline \texttt{vcm} to determine the statistical importance of including the predictor over a model that just includes the simplest multilevel components at the lemma and textual level. As Table 10.7 shows, the model with the predictor reduces the deviance significantly and results in an overall lower \texttt{AIC}. Participle category significantly explains the variability in the response variable, and it is necessary to inspect the model parameters to determine the significance and effect directions of each of the factor levels.

|            | Df  | AIC      | BIC      | logLik | deviance | Chisq  | Chi Df | Pr(>|Chisq|) |
|------------|-----|----------|----------|--------|----------|--------|--------|----------|
| baseline.vcm | 3   | 3139.41  | 3156.77  | -1566.70 | 3133.41  |        |        | \     |
| partcat.rs.text | 21  | 3062.05  | 3183.57  | -1510.02 | 3020.05  | 113.36 | 18     | 0.0000   |

Table 10.7: ANOVA of baseline \texttt{glm} and selected \texttt{glmer} containing participle category information

Table 10.8 reports the parameters of this model. The response variable has been coded such that the logit probability is in terms of \textit{factus est} outcomes, and the response variable is coded such that the level with the largest number of observations on it — namely perfect-verbal — serves as the baseline category to which the remaining four are to be compared. As we can discern, three parameters are significant (excluding the intercept). First, future participles are 0.4311324 times as likely to result in a \textit{factus est} outcome than verbal perfect participles; this is the odds ratio, calculated by exponentializing the logit of the coefficient estimate, $-0.84134$. Gerundive participles are $\exp(-1.72761) = 0.1777086$ times as likely to result in a \textit{factus est} outcome than verbal perfect participles. Third, ambiguous perfect participles are $\exp(-1.47842) = 0.2279976$ times as likely to result in a \textit{factus est} outcome than verbal perfect participles. It is of note that the coefficient estimate for adjectival participles of 0.02 is not significant, because of the variability inherent in the estimate: the standard error is large compared to the coefficient, thereby defining confidence intervals that include zero (they range from $-0.6572124$ to $0.6885334$). The 95% confidence intervals for the three other parameters are firmly in negative logit space, by contrast.

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | 0.44888  | 0.15326 | 2.929   | 0.0034   |
| ParticipleCategoryfuture | -0.84134 | 0.24822 | -3.390  | 0.0007   |
| ParticipleCategorygerundive | -1.72761 | 0.39729 | -4.349  | 1.37e-05 |
| ParticipleCategoryperfect-ambiguous | -1.47842 | 0.22917 | -6.451  | 1.11e-10 |
| ParticipleCategoryadjectival | 0.01566 | 0.34331 | 0.046   | 0.9636   |

Table 10.8: Model parameters of final \texttt{glmer} for serialization $\sim$ participle category

What of text-based differences in the slopes? Figure 10.1 demonstrates clearly that for both future and ambiguous perfect participles the individual by-text slope estimates are all in negative logit space,
10.4 Prepositional Prefixation of the Participle

10.4.1 Introduction

Moving on now to another morphological variable, this section examines whether and to what extent the presence of a prepositional prefix on the participle impacts on the PAC’s serialization. In Latin, as in many other languages, a number of verbs alternate between a morphologically simplex form, e.g. *facio* ‘do’, and one or more prepositional prefixed forms, such as *conficio* ‘complete’ and *perficio* ‘finish’, which comprise the deverbal base *facio* (realized as *ficio*) and the prepositional prefixes *con* and *per*, respectively. The effect of the prefix is typically aspectual, in that it (i) telicizes an atelic predicate or (ii) emphasizes the telicity of an already telic predicate (cf. *morior* ‘die’ vs. *emorior* ‘die, perish’). The prefix can also be spatial, e.g. *gradior* ‘walk’ vs. *progradior* ‘walk forth’.

---

7 See e.g. Devine & Stephens (2013).

8 The vowel change seen in these examples is irrelevant to the discussion. It happened in pre-historic Latin, where an initially placed accent caused the second vowel to weaken: so *cón-facio* > *cón-ficio*.
10.4.2 Effect in AVCs

There has been some discussion in the AVC literature on the effect of different types of prefixation strategies on internal AVC serialization, with the focus being on Germanic. Nothing has been mentioned about this variable’s effect in Latin, though. Let us now report some illustrative examples, from Dutch and German.

In Dutch, participles come with inseparable prefixes, e.g. *ge- in *gemaakt ‘made’, and separable prefixes, e.g. *af- in *afgemaakt ‘finished’. De Sutter (2009) examines the effect this distinction has on (Belgian) Dutch AVC serialization. Fitting a standard glm model to the data, De Sutter (2009: 245) finds that changing the participle from one that has an inseparable prefix to one that has a separable prefix enhances the odds of an auxiliary-first outcome by 3.87 times. While this effect is quite strong, importantly, however, there is no interpretation: it is not clear why this distinction should hold in linguistic terms.

In (early) German varieties, there is also a distinction between inseparable (unstressed) and separable (stressed) prefixes. Sapp (2011) investigates the effect of prefixation on AVCs in Middle High German (MHG) and Early New High German (ENHG). In his MHG dataset (2011: 29–30), he finds that nonfinite forms with a separable prefix, such as *ab gegangen ‘gone away’, prefer the auxiliary-first order with a rate of 40.4% (against a baseline of 28.7%), whereas nonfinite forms with an inseparable prefix, such as *gesegnet ‘blessed’, occur overwhelmingly in the auxiliary-last order with a rate of 76.1% (slightly above the baseline rate of 71.3%). A similar effect is obtained for his ENHG dataset (2011: 64–65). Amongst one of the interpretations Sapp (2011: 65–66) offers is that, given that nonfinite verbs with separable prefixes are typically longer than the nonfinite verbs with inseparable prefixes, the former are placed as late in the clause as possible in accordance with Behaghel’s (1909/1910) Gesetz der wachsenden Glieder. However, there is no statistical assessment with prefixation type pitted against length to corroborate this speculation.

Thus, there is some evidence that prefixation is relevant to serialization choice in AVCs, but most of the evidence at present comes from Germanic varieties, and furthermore the linguistic reasons behind why it is important are unclear and have not been assessed in depth. We will suggest some reasons below, to be probed in other parts of the thesis.

10.4.3 Operationalization

This variable is for the most part straightforward to operationalize by using information in OLD lexical entries. A verb is regarded as having a prefix if (i) a prepositional prefix is listed in its etymological note in OLD, and (ii) the prefixed form has a simplex form with which it alternates. For instance, in the present dataset, we have *affero ‘carry to’, *aufero ‘carry away’, *confero ‘carry together’, *defero ‘carry down’, *differo ‘carry asunder’, *öffero ‘carry out of’, *infero ‘carry in’, *offero ‘carry against’, *perfero ‘carry through’, *praefero ‘carry before’, *profero ‘carry in front’, *refero ‘carry back’, *transfero ‘carry across’, all of which are derived from the simplex verb *fero and the prepositional prefixes *ad, *ab, *con, *de, *dis, *ex, *in, *ob, *per, *prae, *pra, *re and *trans, respectively.

This operationalization, as defined here, excludes the following as prepositional prefixed verbs: *accuso ‘accuse’ is built from *ad+causa+o (there is no verb *cuso, -are); *appello ‘call’ is built from *ad+pello(ere)+o (there is no verb *pello, -are); *condo ‘found’ is built from *con+do, but this *do ‘put’ is a bound morpheme and is only attested in compounds (it is not the free morpheme *do of *dare ‘give’, and is unrelated historically); *conflicto ‘strike down’ is built from *confilio+to (there is no verb *fligo); *congrego ‘assemble’ is built from *con+grex+o (there is no verb *grego); *defendo is built from *de and an unattested verb *fendo; and *extimesco ‘be very frightened’ is built from *ex+timeo+sco (there is no verb *timesco).

The variable is an unordered factor termed PrefixOnParticiple with the level Prefix if the participle has a prepositional prefix on it and the level noPrefix if it does not.

Note that the ‘separable’ prefix is not separated in participle and other non-finite forms, but it is separated in present and simple past tenses.
10.4.4 Statistical Modelling

We turn now to the statistical modelling of this predictor. The relationship of prefixation and PAC serialization choice is presented in Table 10.9. I note a slight discouraging effect of prefixation, in that while participles with no prefix favour the *factus est* serialization with an overwhelming rate of 63.76%, the same is not true for participles with a prepositional prefix, occurring in *factus est* orders with a rate of 55.63%.

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefix</td>
<td>469 (44.37)</td>
<td>588 (55.63)</td>
<td>1057</td>
</tr>
<tr>
<td>noPrefix</td>
<td>490 (36.24)</td>
<td>862 (63.76)</td>
<td>1352</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>959</strong></td>
<td><strong>1450</strong></td>
<td><strong>2409</strong></td>
</tr>
</tbody>
</table>

Table 10.9: Relationship of prefixation and PAC serialization choice

To determine the robustness of this asymmetry, and how it is affected by the multilevel structure inherent in the present dataset, I fit and compare two models. First, a varying intercepts model containing the prefix predictor and varying intercepts for lemma and text is configured:

```r
prefix.ri <- glmer(RealizationOfPAC ~ PrefixOnParticiple + (1|LemmaParticiple) + (1|Text), pac, family = binomial)
```

The second model is the same as the first, with additional terms for varying slopes for only text:

```r
prefix.rs.text <- glmer(RealizationOfPAC ~ PrefixOnParticiple + (PrefixOnParticiple|Text), pac, family = binomial)
```

No models are constructed this time with varying slopes for lemma. This is because verb lemma and prepositional prefix are perfectly correlated (for instance, *inuenio* ‘find’ can only possibly take the level Prefix), and there is no within-group variation to analyse.

A comparison of the models reveals that the more complex model with varying slope structure for text is no more informative than the varying intercepts model (cf. Table 10.10). The latter also exhibits the lower AIC of the two.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>prefix.ri</td>
<td>4</td>
<td>3133.26</td>
<td>3156.41</td>
<td>-1562.63</td>
<td>3125.26</td>
</tr>
<tr>
<td>prefix.rs.text</td>
<td>6</td>
<td>3134.90</td>
<td>3169.62</td>
<td>-1561.45</td>
<td>3122.90</td>
</tr>
</tbody>
</table>

Table 10.10: ANOVA statistics for various glmer containing prepositional prefix information

It is now important to compare the varying intercepts model with the baseline vcm to establish whether the inclusion of the predictor substantially reduces the deviance. The information presented in Table 10.11 shows that it clearly does, reducing it by a value of 8.15, which is obviously significant on a $\chi^2$-distribution with $4 - 3 = 1$ degrees of freedom.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.vcm</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prefix.ri</td>
<td>4</td>
<td>3133.26</td>
<td>3156.41</td>
<td>-1562.63</td>
<td>3125.26</td>
<td>8.15</td>
<td>1</td>
<td>0.0043</td>
</tr>
</tbody>
</table>

Table 10.11: ANOVA of baseline glm and selected glmer containing prepositional prefix information

The parameters of this model are therefore well worth an inspection (cf. Table 10.12). The logit probability is in terms of a *factus est* outcome, and the predictor variable is coded such that the level noPrefix is the baseline. The parameter estimate indicates that when a prefix is added to the participle, the logit probability of a *factus est* outcome decreases by $-0.3051$, which equates to an odds ratio of
0.7370497. Its associated 95% confidence intervals do not span across zero, ranging from \(-0.48543257\) to \(-0.1084965\), and the \(z\)-value is higher than the critical value of 1.96. We therefore have every reason to believe that the “real” population estimate would be negative.

|                  | Estimate | Std. Error | \(z\) value | \(Pr(|z|)\) |
|------------------|----------|------------|--------------|-------------|
| (Intercept)      | 0.3369   | 0.1644     | 2.050        | 0.04040     |
| PrefixOnParticiplePrefix | -0.3051  | 0.1033     | -2.953       | 0.00314     |

Table 10.12: Final glmer for serialization ∼ prepositional prefixation

10.4.5 Discussion

In typological terms, first, it should be noted that the presence of a prefix yields an outcome where the non-finite predicate is placed last in the sequence, as previously found for Belgian Dutch (De Sutter, 2009) and for Middle High German and Early New High German (Sapp, 2011). These typological similarities might be investigated in more depth.

While the effect of this variable is revealing, it is still unclear why it might be important. As Sapp (2011) notes for MHG and ENHG, it may be that the addition of the prefix renders the participle heavier, and so is placed later in the clause in line with Behaghel’s (1909/1910) law. However, Behaghel’s law concerns the relative length of syntactic elements, so it is therefore unclear why the grammar accesses absolute length independently.

Second, we have seen that prefixed participles often have the effect of telicizing an atelic predicate (cf. \(\textit{facio} \ ‘do’ \ vs. \textit{conficio} \ ‘complete’) or focusing on the resultant state of an already inherently telic predict (cf. \(\textit{morior} \ ‘die’ \ vs. \textit{emorior} \ ‘perish’). In approaches of semantics that map the semantics onto the syntax (cf. e.g. Ranchand 2008), the resultant state projection is the lowest in the tree. Thus, predicates with resultant states prefer to stay low, whereas ones without (or ones where the resultant state is not especially salient) are raised higher. We return to this point later.

We could also interpret it in neuro/psycholinguistic terms: the unprefixed participle is semantically less complex, produced and accessed with greater neuromotor fluency, and thus tends to be placed earlier in the clause, whereas the prefixed participle requires greater neuromotor activity, and is placed later.

10.5 Impersonality

10.5.1 Introduction

In Latin, passives formed on transitive verbs, e.g. \(\textit{interfectus est} \ ‘he was killed’ \(\leftarrow\) \textit{interficio} \ ‘kill’, are regularly associated with personal subjects, with the resulting PAC structure inflected for gender, number, and case. The subject can be overtly expressed, as in (19), or elided from the clause, as in (20), an option that Latin has available to it given its status as a canonical example of a pro-drop language.

(19) Lucius Domitius ... ab equitibus \textit{est} Lucius-M.SG.NOM Domitius-M.SG.NOM ... by cavalry-M.PL.ABL be-3SG.PRS.IND

\textit{interfectus}

kill-PTCPL.PRF.M.SG.NOM

‘Lucius Domitius was killed by the cavalry’ (Caes. \textit{Civ. }3. 99. 5)

(20) a militibus nostris \textit{interfecti} sunt by soldier-M.PL.ABL our-M.PL.ABL kill-PTCPL.PRF.M.PL.NOM be-3PL.PRS.IND

‘They were killed by our soldiers’ (B. Hisp. 12. 2)

Passive constructions such as those in (19) and (20) are known in the literature as \textit{personal} passives.
10.5. IMPERSONALITY

Additionally, much as in Dutch, German, Turkish and Welsh and other languages of the world (Siewierska, 1984; Postal, 1986: 9), it is well known that Latin can also form passive constructions from intransitive predicates, e.g. *concursum est* ‘it was run together’ ← *concurro* ‘run together’, as in (21). In these cases, there is no inflected subject, in the sense mentioned above; the “subject” is effectively the event instantiated by the lexical stem of the verb.\(^\text{10}\) The inflectional morphology of the participle is neuter singular. A further example from the verb *pugno* ‘fight’ is given in (22).

(21) ex proximis castellis eo *concursum* est
from nearest-N.PL.ABL fortress N.PL.ABL to.that.place rush-PTCP.PL.PRF.N.SG.NOM be-3SG.PRS.IND
‘They rushed from the nearest fortresses to that place’ (Caes. Gal. 2. 33. 3)

(22) ita uario certamine *pugnatum* est
so changeable-N.SG.ABL contest N.SG.ABL fight-PTCP.PL.PRF.N.SG.NOM be-3SG.PRS.IND
‘So they fought in what was a changeable contest’ (Caes. Civ. 1. 46. 3)

In contrast to the examples in (19)–(20), the PACs in (21)–(22) are *impersonal* passive structures.

10.5.2 Impersonality and PAC serialization

While nothing, to my knowledge, has been mentioned about the effect of impersonality on serialization choice in Dutch and historical German AVCs, it has been raised briefly in connection with PAC serialization. Based on a selection of data from Caesar, Devine & Stephens (2006: 181) assert that impersonal verbs such as that in (21) above, “normally” occur in the *factus est* serialization. For example, the verb *pugno* ‘fight’ “occurs predominantly” in this serialization in the impersonal passive (2006: 188). However, Devine & Stephens (2006) provide no quantitative information, so it is unclear what “normally” and “predominantly” really mean. A quick Caesarean concordance search for *pugnatum* ‘fought’ and the two other participles, *sustentatum* ‘suffered’,\(^\text{11}\) and *concursum* ‘rushed together’, mentioned in their (2006: 181, ex. 79) dataset reveals 16 (84.2%) instances of the *factus est* variant and only 3 (15.8%) of *est factus*.\(^\text{12}\) Devine & Stephens’s observation might therefore be on the right track, but it is obviously not clear how significant this distribution is without full statistics for Caesar’s use of *factus est* vs. *est factus* for impersonal and personal verbs. A detailed quantitative examination of this variable’s effect will therefore be presented in this investigation. Before we do so (in section Section 10.5.4), an operationalization of this input variable is first presented in the next section.

10.5.3 Operationalization

For the purpose of operationalizing impersonal PACs for this study, they are taken to be intransitive verbs that are passivized. The verb is at base (in the active) intransitive (single agentive or single thematic argument verbs) (i.e. does not and cannot take an accusative object) or else verbs that are traditionally regarded as being intransitive (e.g. *uenio* ‘come’ → *uentum est* ‘it was come’, which in the active arguably have an agentive/thematic argument as subject and a goal argument), and when passivized an impersonal structure is formed. The verb is inflected in the neuter singular, as has been documented above and in the later sections on agreement marking. We take intransitives to include verbs such as *utor* which take ablative objects, in that when they are passivized into a gerundive PAC, the ablative object remains ablative. For instance, if we have *usus est gladio* ‘he made use of a sword’, the gerundive passive PAC is

\(^{10}\)Cf. Pinkster (1992: 168) on *fit proelium* ‘a battle took place’, “which may be regarded as a nominalised alternative of the impersonal passive”.

\(^{11}\)Devine & Stephens (2006) treat this as impersonal, which it is flagged as in OLD; the deriving verb *sustento* ‘suffer’ seems to have both transitive and intransitive readings.

\(^{12}\)There are also two examples of discontinuous PACs: *pugnatum... est* (Caes. Gal. 2. 33. 3) and *esse... pugnatum* (Caes. Civ. 1. 46. 1).
utendum est gladio (ei) ‘it must be used with a sword (by him)’ rather than a canonical passive structure
*utendus est gladius (ab eo) ‘the sword must be used by him’. These verbs were accordingly tagged with
the level impersonal, and all others — like those in (19) and (20) above — were tagged with the level
personal.

10.5.4 Statistical modelling

I note immediately that (i) there are comparatively few observations with the level impersonal and that
(ii) there is a slight apparent asymmetry between the distribution of impersonal vs. personal verbs with
respect to serialization choice, which is in the direction suggested by Devine & Stephens (2006). However,
this asymmetry is not statistically significant on a basic \( \chi^2 \)-test of independence (\( \chi^2 = 2.2719, df =
1, p = 0.1317 \)). We will now explore this relationship in more detail via statistical models.

<table>
<thead>
<tr>
<th>Serialization</th>
<th>factus (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>personal</td>
<td>924 (40.16)</td>
<td>1377 (59.84)</td>
<td>2301</td>
</tr>
<tr>
<td>impersonal</td>
<td>35 (32.41)</td>
<td>73 (67.59)</td>
<td>108</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
<td>2409</td>
</tr>
</tbody>
</table>

Table 10.13: Relationship between impersonality and PAC serialization

Unlike prefixation, there is within-lemma variability for impersonality: the same lemma may have
personal or impersonal realizations, depending on choices made by the speaker. For instance, *pugno ‘fight’
typically has the impersonal passive realization *pugnatum est ‘it was fought’, but the personal passive
realization can also be found, e.g. *pugna pugnata est ‘a fight was fought’. Similarly, *proficiscor ‘depart’
is a personal deponent structure with the perfect participle, but with the gerundive it is an impersonal
passive structure *proficiscendum est ‘it is needing to be departed’. Given this potential within-lemma
variability, the usual four models can be constructed and compared. First, a varying intercepts model
containing the predictor and varying intercepts for lemma and text is configured:

```
impersonal.ri<-glmer(RealizationOfPAC~Impersonality+(1|LemmaParticiple)+(1|Text),pac,family=binomial)
```

The second model is the same as the first, but with the addition of varying slopes for lemma and text.

```
impersonal.rs.both<-glmer(RealizationOfPAC~Impersonality+(Impersonality|LemmaParticiple)+(Impersonality|Text),pac,family=binomial)
```

The third model contains varying slopes for only lemma:

```
impersonal.rs.lemma<-glmer(RealizationOfPAC~Impersonality+(Impersonality|LemmaParticiple)+(1|Text),pac,family=binomial)
```

And the fourth model contains varying slopes for only text:

```
impersonal.rs.text<-glmer(RealizationOfPAC~Impersonality+(1|LemmaParticiple)+(Impersonality|Text),pac,family=binomial)
```

The four models are compared in terms of their degrees of freedom, \( AIC, BIC, \) log likelihood, and
deviance in Table 10.14. This table clearly shows that the addition of complex random effect structure
— that is, allowing the slopes to vary between texts and/or lemmas — hardly results in a better model
fit (the deviances of the varying slope models are only fractionally smaller than that for the varying
intercept only model). Second, the \( AIC \) of impersonal.ri is the closest to zero, meaning that it should
be preferred for further appraisal.
10.6. DEPONENCY

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>impersonal.ri</td>
<td>4</td>
<td>3139.24</td>
<td>3162.39</td>
<td>-1565.62</td>
<td>3131.24</td>
</tr>
<tr>
<td>impersonal.rs.lemma</td>
<td>6</td>
<td>3142.87</td>
<td>3177.59</td>
<td>-1565.43</td>
<td>3130.87</td>
</tr>
<tr>
<td>impersonal.rs.text</td>
<td>6</td>
<td>3142.98</td>
<td>3177.70</td>
<td>-1565.49</td>
<td>3130.98</td>
</tr>
<tr>
<td>impersonal.rs.both</td>
<td>8</td>
<td>3146.78</td>
<td>3193.07</td>
<td>-1565.39</td>
<td>3130.78</td>
</tr>
</tbody>
</table>

Table 10.14: ANOVA statistics for various GLMERS containing impersonality information

The varying intercepts only model is subsequently compared with the baseline varying components model by means of ANOVA, displayed in Table 10.15. This indicates that there is very little to be gained by adding the predictor of impersonality: the $-2LL$ reduction of 2.17 is not significant on a $\chi^2$-distribution with $4 - 3$ degrees of freedom ($p = 0.141$), and the AIC is virtually unchanged. Consequently, there are no parameters that are worthy of inspection. The best model for the impersonality predictor is the baseline vcm, that is the one that contains only information at the lemmatic and textual levels.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.vcm</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>impersonal.ri</td>
<td>4</td>
<td>3139.24</td>
<td>3162.39</td>
<td>-1565.62</td>
<td>3131.24</td>
<td>2.17</td>
<td>1</td>
<td>0.1408</td>
</tr>
</tbody>
</table>

Table 10.15: ANOVA of baseline GLM and selected GLMER containing impersonality information

In sum, impersonality cannot be considered an influencer of serialization choice. The statistical model therefore confirms what we saw above in the very basic cross-tabulation in Table 10.13.

10.5.5 Discussion

Having constructed the above models, we can see that impersonality does not have an effect on serialization choice in PACS. But why, especially given that Devine & Stephens (2006) mention it as a potential source of influence? It is important to note at this point that there is reason to believe that Devine & Stephens (2006) do not regard impersonality as a primary conditioning factor per se. Indeed, they attribute this preference for the factus est serialization in impersonal passives to the latter’s eventive characteristics (2006: 181, 188; cf. 158), and with the lack of event “participants that could be focused”. I note that such characteristics of impersonals are implicit in Colebourn’s (1948: 81, n.2) assertion that “[t]he idea in an impersonal passive is to focus the interest on the action, the doer being too vague or too obvious to mention [italics in original]” and in Pinkster 1992: 169f., who asserts that “the event is presented not from the perspective of one of the participants, but as such [i.e. from the perspective of the event]. A clause with an impersonal passive is a statement about what happened rather than about who did what. We might call this ‘promotion’ of the action involved”. Such claims suggest that the impersonal passive might only stimulate factus est by virtue of its inherent association with other potentially more influential factors, such as (i) eventivity of the clause, (ii) focus on the action, (iii) lack of focus on participants and/or (iv) lack of participants more generally. The results presented here support the idea that impersonality is not important in and of itself.

10.6 Deponency

10.6.1 Introduction

A set of Latin verbs are characterized as being deponent. This means that they take passive morphology, such as the -r passive in one-word verb forms and PACS in perfect tenses, but they have ‘laid aside’ their

13Though, they do not explicitly say this. This is my interpretation of what they say.
14In panchronic terms, it has been claimed that there are 884 “déponents proprement dits”, i.e. true deponents (Flobert, 1975).
passive meaning and are consequently active in their interpretation. *Proficiscor* ‘set out, depart’ and *sequor* ‘follow’ are two well-attested examples of this type. So, while *secatus est* is passive in form, it is active in meaning: it means ‘he followed’ rather than ‘he was followed’. By contrast, other verbs (the majority) are not deponent. If their form is passive, they retain their passive interpretation. *Amo* ‘love’ and *paro* ‘prepare’ are two examples of this latter type.

### 10.6.2 Deponency and PAC Serialization

This difference between deponents and what I shall term non-deponents for the purposes of this discussion has been briefly brought up in connection with the PAC alternation. Spevak (2006, 2010) provides tables showing how serialization choice is distributed amongst deponent verbs on the one hand (2006: 277, Table 4; 2010: 155, Table 14) and amongst non-deponent verbs, or “passifs, passives” in her terminology, on the other (2006: 272, Table 3; 2010: 151, Table 13). I collapse and crosstabulate this information in Table 10.16.15

<table>
<thead>
<tr>
<th></th>
<th>factus est (%)</th>
<th>est factus (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>deponent</td>
<td>70 (78.65)</td>
<td>19 (21.35)</td>
<td>89</td>
</tr>
<tr>
<td>non-deponent</td>
<td>150 (73.17)</td>
<td>55 (26.83)</td>
<td>205</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>220</strong></td>
<td><strong>74</strong></td>
<td><strong>294</strong></td>
</tr>
</tbody>
</table>

Table 10.16: The relationship between deponency and PAC serialization according to Spevak (2006, 2010)

On the basis of these statistics, Spevak claims that the *factus est* variant is “dominant” in deponent verbs, and “even more so than in the case of the passive voice” (2010: 155). The *est factus* serialization, by contrast, is used “less frequently”. The wording is suggestive of the existence of a relationship. But Spevak did not use any type of statistical test. If we conduct a $\chi^2$-test of independence over Spevak’s data, we obtain a non-significant result and a low effect size ($\chi^2 = 0.7202, df = 1, p = 0.3961, CI = [0.7236727, 2.5964431], OR = 1.349535, \phi = 0.058$). Consequently, the existence of a relationship is not borne out, and the distribution we see in Table 10.16 could well be down to random fluctuations.

### 10.6.3 Operationalization

To operationalize deponency, I draw on the list of Flobert (1975), which is still the source of choice when it comes to Latin deponent verbs. In Flobert’s (1975: 664–681) list a distinction is made between true deponents (“déponents proprement dits”), deponent variants (verbs that have deponent variants), neo-deponents (verbs undergoing deponentization), and semi-deponents. For the purposes of the present study, I shall combine these under the level deponent.16 There are several reasons for this. First, there are few observations with semi-deponent predicates ($n = 45$) and for variable and neo-deponents ($n = 23$). Second, an initial $\chi^2$-test revealed there may be something informative going on with this variable, as all contributions to $\chi^2$ revealed a significant $p$-value ($\chi^2 = 10.2667, df = 4, p = 0.03617$),17 thereby allowing us to proceed with pairwise comparisions. A pairwise test of proportions with Holm’s adjustment revealed that pairwise differences between deponents and other types of deponency were not significant ($p = 1$).18

---

15Actually, this table suppresses data Spevak also reports on the other realizational variants, i.e. *factus ... est* and *est ... factus.* It did not seem relevant to include them in the table here.

16While it is not listed as deponent in Flobert’s list, I also include *fio* ‘become’ in perfect participle forms, where it is structured in a PAC utilizing the lemma *facio* ‘make’, thus akin to verbs such as *soleo* ‘be accustomed’ and *audeo* ‘dare’.

17The original levels being: non-deponent, other-deponent, semi-deponent, and fio/faco (see Footnote 16).

18Neither were there significant differences between non-deponents and the other types of deponency, but there is little theoretical motivation for combining these levels.

Another statistical possibility would be to run the regression models excluding other types of deponency, that is with just true deponents vs. non-deponents. However, if we do this, there is still very little to be gained empirically, as the regression parameters remain similar for both the standard GLM ($\beta = 0.2649, SE = 0.1043, z = 2.540, p = 0.0111$) and the multilevel GLM ($\beta = 0.2582, SE = 0.1076, z = 2.399, p = 0.0164$). Effectively, then, we are only reducing the sample size from $n = 2409$ to $n = 2201$ by omitting these levels from the analysis.
10.6.4 Bivariate modelling

Table 10.17 presents a crosstabulation of the relationship between deponency and PAC serialization. What this shows is that, while 64.97% of deponent verbs are realized with *factus est* PACs, 58.14% of non-deponents are so realized. This very slight asymmetry we see here, which is significant on a $\chi^2$-test of independence ($\chi^2 = 9.5841$, $df = 1$, $p = 0.001963$), is reminiscent of Spevak’s (2006; 2010) data presented above, although the overall share of *factus est* is smaller in the present dataset because of text sampling decisions. (Recall that in our dataset the overall share of *factus est* is 60.15%, while in Spevak’s dataset its share is 74.83%.)

<table>
<thead>
<tr>
<th></th>
<th>Serialization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>est factus</em> (%)</td>
</tr>
<tr>
<td>non-deponent</td>
<td>705 (41.86)</td>
</tr>
<tr>
<td>deponent</td>
<td>254 (35.03)</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>959</strong></td>
</tr>
</tbody>
</table>

Table 10.17: Relationship between deponency and PAC serialization

However, what is important is whether deponents are significantly different from non-deponents with respect to serialization choice in PACs when we take into account the multilevel structure inherent in the present dataset. This time two multilevel models are constructed and compared. The first contains the deponency predictor and random intercepts for the verb lemma and text:

```r
deponent.ri<-glmer(RealizationOfPAC~Deponency+(1|LemmaParticiple)+(1|Text),pac,family=binomial)
```

The second model is the same as the first, but with the addition of varying slopes for the text.

```r
dpeonent.rs.text<-glmer(RealizationOfPAC~Deponency+(1|LemmaParticiple)+(Deponency|Text),pac,family=binomial)
```

Further models that contain information on varying slopes for the verb lemma are not included because the predictor’s level, at least as coded here, is uniquely identifiable by the lemma, and there is consequently no within-group variability that could influence the outcome.

An ANOVA comparison of the varying intercept only model `deponent.ri` with that including varying slopes for text `deponent.rs.text` reveals that there are no influential interactions between the predictor and the textual source (as demonstrated in Table 10.18): texts behave more or less the same when it comes to deponency.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>deponent.ri</td>
<td>4</td>
<td>3136.14</td>
<td>3159.29</td>
<td>-1564.07</td>
<td>3128.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>deponent.rs.text</td>
<td>6</td>
<td>3139.00</td>
<td>3173.72</td>
<td>-1563.50</td>
<td>3127.00</td>
<td>1.15</td>
<td>2</td>
<td>0.5631</td>
</tr>
</tbody>
</table>

Table 10.18: ANOVA statistics for various GLMERs containing deponency information

Secondly, the varying intercepts model is compared with the baseline VCM (cf. Table 10.25). The ANOVA comparison indicates that the inclusion of the predictor substantially improves the model fit, as the deviance is reduced by 5.26 significant on a $\chi^2$-distribution with $4 - 3 = 1$ degrees of freedom, and its $AIC$ is lower.\(^{19}\)

The inspection of the parameters of the model `deponent.ri` is therefore warranted. Table 10.20 reports the model, wherein the logit probability is in terms of a *factus est* outcome, and the predictor variable is dummy coded such that *non-deponent* is the baseline level. The second line in Table 10.20 demonstrates that, when the PAC is composed of a deponent verb as opposed to a non-deponent verb,\(^{19}\)

\(^{19}\)The alternative $BIC$ measure is not lower, however, but I use $AIC$ throughout for consistency.
the logit probability of a *factus est* increases by 0.2959, which corresponds to an odds ratio of 1.344336. The 95% confidence intervals do not cross zero (from 0.04651126 to 0.5452300) and the absolute z-value is above 1.96, with the result that we can be confident that the population slope lies in positive logit space between those intervals.

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | 0.1078 | 0.1592 | 0.677 | 0.498 |
| Deponency | 0.2959 | 0.1272 | 2.325 | 0.020 |

Table 10.20: Final glmer for serialization ~ deponency

10.6.5 Discussion

It is not particularly clear to me why deponency has such an effect on serialization choice in PACS. One possibility is that it is not entirely independent of another morphological contrast — namely, voice (e.g. active vs. passive). Voice is more interpretable and predictive in cognitive terms: the participle that expresses the morphosyntactic property of being active is put first because it is cognitively more highly accessible than passives, being easier to produce and process. In short, deponency might be of ancillary importance, but only through its connection with voice. Be that as it may, there is good empirical evidence to dismiss this potential epiphenomenality. A crosstabulation of the relationship between grammatical voice and serialization choice is presented in Table 10.21. We can see that the distribution is similar to that of deponency, with actives preferring *factus est* slightly more strongly than *est factus*.

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>active</td>
<td>253 (36.04)</td>
<td>449 (63.96)</td>
<td>702</td>
</tr>
<tr>
<td>passive</td>
<td>706 (41.36)</td>
<td>1001 (58.64)</td>
<td>1707</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
<td>2409</td>
</tr>
</tbody>
</table>

Table 10.21: Relationship between voice and PAC serialization

While a standard glm returns this predictor as a salient influence ($\beta = -0.22691, SE = 0.09270, z = -2.448, p = 0.0144$; treatment coding used, where active = 0, passive = 1), a hierarchical model is not as supportive, with it there having only a marginally significant effect ($\beta = -0.16709, SE = 0.09544, z = -1.751, p = 0.0800$). Consequently, it appears that voice is not the reason that deponent verbs opt for *factus est* serializations proportionally more often than non-deponent verbs.\(^{20}\)

\(^{20}\)Ideally, one would include both voice and deponency in a glm to test their joint effect. However, this is problematic here, as the two predictors are highly collinear. First, a glm with both predictors produces a Variance Inflation Factor (VIF) score 2.07 (values > 2 indicate that multicollinearity may be present). In this glm neither variable is important, another indication of collinearity. Furthermore, when voice is modelled as a function of deponency, the latter is strongly predictive, where Nagelkerke’s $R^2$ equals 0.797 and $C$ equals 0.922. The notion of multicollinearity is discussed later in relation to multivariate modelling strategies.
10.7 Conjugational Class

10.7.1 Introduction

Latin grammarians divide verbs into conjugational classes depending on their morphophonological stem properties of the infectum. We learn, for instance, that the first conjugation is formed on "a"-stems (e.g. amo, -are, amavi ‘love’), the second conjugation is formed on "e"-stems (e.g. moneo, -ere, monui ‘warn’), the third conjugation is formed on consonant stems (e.g. mitto, -ere, missi ‘send’), and the fourth conjugation is formed on "i"-stems (e.g. audio, -ire, audi ‘hear’).

Some further typical members of the four classes are given in Table 10.22.

<table>
<thead>
<tr>
<th>Conjugational Class</th>
<th>Example</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>alien-o, -a-re</td>
<td>‘alienate’</td>
</tr>
<tr>
<td></td>
<td>demonstr-o, -a-re</td>
<td>‘show’</td>
</tr>
<tr>
<td></td>
<td>labor-o, -a-re</td>
<td>‘work’</td>
</tr>
<tr>
<td></td>
<td>pugn-o, -a-re</td>
<td>‘fight’</td>
</tr>
<tr>
<td>Second</td>
<td>caue-o, -e-re</td>
<td>‘beware of’</td>
</tr>
<tr>
<td></td>
<td>persuade-o, -e-re</td>
<td>‘persuade’</td>
</tr>
<tr>
<td></td>
<td>remoue-o, -e-re</td>
<td>‘remove’</td>
</tr>
<tr>
<td></td>
<td>retine-o, -e-re</td>
<td>‘hold back’</td>
</tr>
<tr>
<td>Third</td>
<td>accip-i-o, -e-re</td>
<td>‘receive’</td>
</tr>
<tr>
<td></td>
<td>colloqu-or, -i</td>
<td>‘converse’</td>
</tr>
<tr>
<td></td>
<td>obser-o, -e-re</td>
<td>‘plant’</td>
</tr>
<tr>
<td></td>
<td>tang-o, -e-re</td>
<td>‘touch’</td>
</tr>
<tr>
<td>Fourth</td>
<td>exori-or, -i-re</td>
<td>‘rise’</td>
</tr>
<tr>
<td></td>
<td>poli-or, -i-ri</td>
<td>‘obtain’</td>
</tr>
<tr>
<td></td>
<td>reperi-o, -i-re</td>
<td>‘find’</td>
</tr>
<tr>
<td></td>
<td>ueni-o, i-re</td>
<td>‘come’</td>
</tr>
</tbody>
</table>

Table 10.22: Latin Conjugational Classes

10.7.2 Effect in AVCs

To my knowledge there has been no discussion of the effect of conjugational class on AVC or Latin PAC serialization. I hope that the present section will enhance our understanding in this area.

10.7.3 Variable profile and operationalization

The operationalization of this variable is largely straightforward, by using conjugational class information under verb lemmas in Lewis & Short (1897), available online via http://www.perseus.tufts.edu/. Some verbs, however, are not assigned to conjugational classes, because they are thought to be irregular (they show suppletion in the paradigm, for instance): these include sum ‘be’, fero ‘carry’ and eo ‘go’. As a result, we have an unordered factor with the levels 1, 2, 3, 4, and irregular. An overall $\chi^2$ test revealed that the serialization choice is not independent of conjugational class ($\chi^2 = 10.3405, d.f. = 4, p = 0.03507$), but given the number of levels ($k = 5$), I decided to iteratively collapse them via pairwise comparisons using Holm’s correction. As a result, I contrast those verbs that have a vowel stem in the infectum vowelStem (i.e. original levels 1, 2, and 4) versus those verbs that have a consonant or athematic stem in the infectum consonantStem (i.e. original levels 3 and irregular), thereby reducing this unordered factor to a binary one.

---

21 This characterisation is not quite accurate, but it will do for present purposes. For more details on conjugational class formation, one might consult Palmer (1954: 266-270).

22 I give 1.sg.ind and infinitive forms in Table 10.22.
10.7.4 Statistical modelling

This section discusses the bivariate modelling of this variable. Table 10.23 shows the relationship of conjugational class and PAC serialization. We observe a slight asymmetry here, with verbs with vowel stems (i.e. conjugational classes 1, 2, and 4) associating more overwhelmingly with factus est serializations than verbs with consonant stems (i.e. conjugational class 3 and verbs with irregular, athematic paradigms). How strong is this effect, though, in statistical modelling terms?

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>consonant stem</td>
<td>608 (42.46)</td>
<td>824 (57.54)</td>
<td>1432</td>
</tr>
<tr>
<td>vowel stem</td>
<td>351 (35.93)</td>
<td>626 (64.07)</td>
<td>977</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
<td>2409</td>
</tr>
</tbody>
</table>

Table 10.23: Relationship of conjugational class and PAC serialization

As with prefixation and deponency, conjugational class exhibits no variation with respect to verb lemma: porto ‘carry’ is always first conjugation, audio is always fourth conjugation, and so on. Only two models can be constructed, therefore. The first model includes an input term for the predictor and random intercepts for lemma and text:

```r
conjugation.ri <- glmer(RealizationOfPAC~ConjugationalClass+(1|LemmaParticiple)+(1|Text), pac, family=binomial)
```

The second model is the same as the first, but with the addition of varying slopes for the text.

```r
conjugation.rs.text <- glmer(RealizationOfPAC~ConjugationalClass+(1|LemmaParticiple)+(ConjugationalClass|Text), pac, family=binomial)
```

The resulting ANOVA comparison between the two models — reported in Table 10.24 — demonstrates that the inclusion of the varying slopes for text significantly enhances the model in terms of its reduction in deviance (p = 0.0357). The more complex model should be preferred. Note additionally that its AIC is lower.

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>conjugation.ri</td>
<td>4</td>
<td>3133.68</td>
<td>3156.82</td>
<td>-1562.84</td>
<td>3125.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>conjugation.rs.text</td>
<td>6</td>
<td>3131.01</td>
<td>3165.73</td>
<td>-1559.51</td>
<td>3119.01</td>
<td>6.66</td>
<td>0.0357</td>
</tr>
</tbody>
</table>

Table 10.24: ANOVA statistics for various glmer models containing conjugational class information

Secondly, conjugation.rs.text is to be compared with the baseline vcm to determine whether the inclusion of the predictor, whether on its own (the fixed effect slope) or via its interaction with text (the varying slope term), significantly decreases the deviance. The ANOVA output in Table 10.25 shows that it does, decreasing it by 14.39 which is significant on 6 − 3 = 3 degrees of freedom (p = 0.0024).

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.vcm</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>conjugation.rs.text</td>
<td>6</td>
<td>3131.01</td>
<td>3165.73</td>
<td>-1559.51</td>
<td>3119.01</td>
<td>14.39</td>
<td>0.0024</td>
</tr>
</tbody>
</table>

Table 10.25: ANOVA of baseline glm and selected glmer containing conjugational class information

The parameters of conjugation.rs.text need to be inspected. Here, the variable RealizationOfPAC is the response variable with the level for est factus the baseline (i.e. coded 0) and the level for factus est treated as a success (i.e. coded 1), and the variable ConjugationalClass is the input. For this predictor variable, treatment (0-1) coding is used, with the level consonantStem configured as the baseline (i.e. coded 0) and the level vowelStem as a success (i.e. coded 1). The resulting model is reported in
10.7. CONJUGATIONAL CLASS

Figure 10.2: Individual by-text coefficients for glmer containing conjugation class information

Table 10.26: The second line in the table indicates that verbs whose infectum form is of a vowel stem encourage factus est serializations more than those verbs which are not. Specifically, compared to when the verb’s infectum form has a consonant/athematic stem, when the verb’s infectum form has a vowel stem the logit probability of factus est is increased by 0.36750, corresponding to an odds ratio of 1.44412. The 95% confidence intervals lie between 0.01412567 and 0.7208729, which are quite wide but nonetheless completely in positive logit space. The estimate divided by its standard error results in a z-value of 2.038, which is greater than the critical value of 1.96. It is therefore a significant parameter estimate, acting robustly on the alternation.

|                          | Estimate | Std. Error | z value | Pr(>|z|) |
|--------------------------|----------|------------|---------|----------|
| (Intercept)              | 0.02881  | 0.17418    | 0.165   | 0.8686   |
| ConjugationalClassvowelStem | 0.36750  | 0.18030    | 2.038   | 0.0415   |

Table 10.26: Final glmer for serialization ∼ conjugational class

The statistical soundness of the effect can be discerned from Figure 10.2. This shows that all the individual by-text coefficient slopes are firmly in positive logit space, except for in Varro’s Res Rusticae.

10.7.5 Discussion

Why should conjugational class be important? One possibility, as we have seen for a number of other variables, is that conjugational classes 1, 2, and 4 are easier to access and produce compared with conjugational class 3 and irregulars. This would make sense, as participial morphemes (such as -tus and -turus) are simply “tacked on” to the stem in the former group with minimal phonological changes (e.g. porta- → portatus), whereas for conjugations with consonantal stems more complex phonological changes are necessary (e.g. fingo → fictus); indeed, for irregular forms, the participle may have to be accessed from a completely different source than the lemma (e.g. the future participle for sum is futurus). Fronting the auxiliary may “buy more time” for such participles to be accessed/formed in the planning scheme of
the sentence.
10.8 Number

This section discusses the potential for the grammatical category of number to exert an influence on the PAC alternation. Most basically, we will be concerned with the distinction between singular, which expresses a single entity, and plural, which expresses more than one entity, and we will be concerned exclusively with number as encoded on nominals and which spreads to the verb phrase by means of agreement rules.

10.8.1 Linguistic importance of number

10.8.1.1 General background

Markedness accounts argue that the singular is “unmarked” whereas the plural is “marked” (Greenberg, 1966; Eberhard, 1997; Corbett, 2000; Croft, 2003). There is reliable evidence for this claim, in that singular and plural show asymmetrical distributions in language. We survey some of the evidence below.

The unmarked category, i.e. singular, is often zero realized in that there is no morphophonemic indication of its number value, whereas the marked category, i.e. plural, often has (Greenberg, 1966; Payne, 1997; Croft, 2003). In other words, it’s typical to get languages that express (i) plural marking overtly but no singular marking, (ii) both singular and plural marked overtly, or (iii) neither are marked. It’s typologically rarer to find languages that mark singular, but not the plural (Croft, 2003: 89). This information is captured in Croft’s (2003: 89) table, reported here as Table 10.27. In English, for instance, the singular is unmarked, e.g. dog, whereas the plural is overtly indicated by means of -(e)s, e.g dogs.

<table>
<thead>
<tr>
<th>No singular inflection</th>
<th>Overt plural inflection</th>
<th>No plural inflection</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 10.27: Marking of number values (Croft, 2003: 89)

Greenberg (1966); Corbett (2000); Croft (2003) and many others discuss the notion of “syncretization” (Greenberg, 1966), “inflectional potential” (Croft, 2003), or “intracategory variability” (Haskell et al., 2011). All this means is that the unmarked category will have more morphological forms associated with it than the marked category. Put more formally, “if the marked value has a certain number of formal distinctions in an inflectional paradigm, then then unmarked value will have at least as many formal distinctions in the same paradigm” (Croft, 2003: 97). Many examples are mentioned in the literature concerning the category of number. In the grammar of Ancient Greek, nominative and vocative cases are distinguished in the singular, but in the plural the distinction is neutralized and both have the same form (Croft, 2003: 27). In English third person pronouns, there is “a clear asymmetry” between the distribution of singular and plural forms. There are three forms in the singular he, she, it, but in the plural only one they (Croft, 2003). In Russian past tense verb paradigms, a three-way distinction in gender is made in the singular, but there is no distinction whatsoever in the plural: for instance, byl ‘was’ is distinguished in the singular byl (masc.), byl-a (fem.), byl-o (neut.), whereas in the plural the form is byl-i for all three genders (Corbett, 2000: 272).

Unmarked forms have wider semantic application. The singular form is the neutral form and can be used to express either the singular or the plural, whereas the plural form can indicate only ‘more than one’ entity. This type of double function is termed “facultative expression” by Greenberg (1966: 28). An

---

23 Languages have other levels of number, such as dual and paucal (e.g. Corbett 2000).

24 There is additionally the notion of verbal number or pluractionality, which will not concern us here. It has much to do with aspect and Aktionsart, which I deal with later on.

25 Semantic accounts of number, however, contend that singular is the more marked category (Andrew Koontz-Garboden, p.c.).
example is cited of the Korean number system where the singular, which is zero-marked, can be used to indicate a single entity (i.e. used as singular) or more than one entity (i.e. used as plural); by contrast, the plural marker -tul, when it is used, can only be applied to nouns that specify more than one entity.

Unmarked forms have been claimed to be more frequent in texts than marked forms, and this seems to be true of the singular vs. plural distinction (e.g., Greenberg 1966; Corbett 2000; Croft 2003). Corbett (2000: 281) offers data from a number of languages, pointing to the same asymmetry. Table 10.28 gives the relevant statistics, showing that — generally — singular forms occur between 64% and 85.2% of the time, much more than often the plurals.

These typological differences have psycholinguistic correlates. For example, Jakobson (1968) notes that aphasic patients typically substitute marked forms with unmarked forms, and that marked forms of categories are usually the first to be lost from the grammar of such patients. Studies are cited in Ulatowska & Baker (1975), showing evidence of this in the case of grammatical number. Further psycholinguistic evidence comes from agreement mismatches in subject-verb strings, where the subject NP comprises a head noun and a preverbal noun, e.g. The key to the cabinet where key is the head noun and cabinet is the preverbal (or: local) noun. Bock & Miller (1991); Eberhard (1997) have found significantly more mismatches when the head noun is singular and the preverbal noun is plural, e.g. The key to the cabinets [were]... than when the head noun is plural and the preverbal noun is singular, e.g The keys to the cabinet [was]... Eberhard (1997) argues for an activation based account. Here, the plural is specifically marked with a number feature, [+plural], while the singular is not, [∅], the latter being the “default” number value. The evidence for this comes from both observational and experimental research.

All of the above evidence points to a strong asymmetry in the processing and production of the differing levels of number, which may be interpreted in terms of markedness if so desired.

### 10.8.1.2 Effects of number on variable grammatical realization

This asymmetry of singular and plural number just mentioned extends into the domain of variable grammatical realization. For instance, Rohdenburg (2003: 223–225) looks at the overtness of prepositional links between a matrix NP and a dependent interrogative clause as in inquiries [∅/of] what happened as a function of the grammatical number of the matrix NP. As we have just seen, it has been claimed that plural nouns are cognitively more complex to access and produce than singular nouns. Rohdenburg (2003: 205) argues that a “complexity principle” governs grammar, in that “in the case of more or less explicit constructional options ... the more explicit one(s) will tend to be preferred in cognitively more complex environments”. Consequently, he hypothesizes that plural matrix NPs, because of their higher degree of complexity, will tend to introduce the more grammatically explicit construction with the preposition realized, whereas singular matrix NPs, because of their lower degree of complexity, will tend to introduce the less explicit variant in which the preposition is omitted. This hypothesis is first supported with inquiry/ies on British English data. I conducted statistical tests over Rohdenberg’s data and found the results to be significant $p < 0.05$ (specifically, $\chi^2 = 5.4, df = 1, p = 0.02, \phi = 0.27$). The hypothesis is additionally supported with indication/s on British English data: $\chi^2 = 12.14, df = 1, p < 0.001, \phi = 0.12$. 

### Table 10.28: Frequency of singular nominal forms (Corbett, 2000)

<table>
<thead>
<tr>
<th>Language</th>
<th>% Singular</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>74.3</td>
</tr>
<tr>
<td>Latin</td>
<td>85.2</td>
</tr>
<tr>
<td>Russian (Josselson)</td>
<td>77.7</td>
</tr>
<tr>
<td>Russian (Stejnfeldt)</td>
<td>71.5</td>
</tr>
<tr>
<td>Sanskrit</td>
<td>70.3</td>
</tr>
<tr>
<td>Slovene (Oz.)</td>
<td>72.5</td>
</tr>
<tr>
<td>Slovene (New.)</td>
<td>75.3</td>
</tr>
<tr>
<td>Upper Sorbian</td>
<td>64</td>
</tr>
</tbody>
</table>

174
Note that the effect with indication/s is slightly weaker than with inquiry/ies, $\phi = 0.27$ as compared with $\phi = 0.12$ respectively. This suggests that lexical differences may additionally be involved, pointing to a multivariate effect. Rohdenburg (2002), cited in the above, finds similar effects with the realization of prepositions that introduce gerundial clauses, in that the type of realization is taken as a function of the number of the matrix NP. Specifically, it is found that plural NPs are more likely to introduce gerund clauses with the preposition overtly realized, e.g. He had difficulties in finishing his paper, whereas singular NPs are more likely to introduce gerund clauses with the preposition elided, e.g. He had difficulty finishing his paper.

Bresnan et al. (2007: 77) observe that “[f]rom cross-linguistic evidence, number (singular/plural) . . . could also have an influence” on the choice of construction in the English dative alternation. Accordingly, they configure variables for the number of recipient and number of theme. Estimating the probability of a to-dative outcome with hierarchical regression models, they find that, as compared to singular themes, plural themes significantly increase the odds of a prepositional dative outcome by 1.65 times (in their Model A). The number of recipient is not a significant model parameter. Bresnan & Ford (2010) likewise include number as a variable in their corpus model of the dative alternation, but here it is found to be a marginally significant parameter ($p = 0.0805$).

### 10.8.1.3 Grammatical number in AVCS and in Latin PACS

#### 10.8.1.3.1 Morphological encoding

It is well-known that, cross-linguistically, verbs are syntactic loci of agreement relations (e.g. Corbett 2000; Baker 2008). Verbs typically agree with the subject argument. For example, it is of some importance in the grammar of English, with third person present verbs encoded for number of the grammatical subject — contrast singular presents which are marked with -s, e.g. he come-s, with third person plural presents which are unmarked, e.g. they come.

Given the predicational nature of AVCS, we find that these constructions are often coded for the nominal number of their arguments. There are several candidates for the locus of number agreement information in such verb clusters: drawing on terminology from Anderson (2006), it can be located on the auxiliary element (“auxiliary-headed”), on the lexical element (“lexical-headed”), or on both (“double-headed”). A simple example of an AVCS where number is “auxiliary-headed” would be the English he is singing vs. they are singing: the auxiliary is marked for number, but the non-finite verb (the participle) is not. We will use this terminology below.

With regard to the grammar of Latin, a distinction is made between singular and plural number values. It is expressed morphologically in the inflectional ending of a word — thus, seru-us ‘slave’ vs. seru-i ‘slaves’. Latin is a strongly dependent marking language, making use of syntactic means to express number by agreement rules between the subject argument and the verb, a familiar pattern in Indo-European languages persisting in Romance varieties. Consequently, Latin PACS are no exception to the general tendency for AVCS to mark argument number. However, there are some important differences depending on the degree of finiteness and impersonality of the verbal predicate and the constituency of the subject argument. We look at these in more detail.

In finite personal PACS, the grammatical number (singular vs. plural) of the overt or covert subject spreads to the verbal predicate, and is thus marked on both the auxiliary and the participle: to adopt the terminology of Anderson (2006) it is double-headed. To illustrate, in (23a) the subject phrase omnisque noster equitatus is explicitly morphologically singular, and this information is carried over on to the auxiliary’s form est and to the ending of the participle -us via agreement rules. By contrast, in (23b), the plural quantifier subject plures triggers the PAC to be plurally marked — this is shown by the plural form of the auxiliary sunt and the ending -i of the participle.  

---

26There are plenty of examples in Anderson (2006).
27While there are some vestiges of dual marking in e.g. ambo ‘both’ and duo ‘two’ this does not mean that Latin has a dual value in the same way that Homeric Greek does, for instance, where it is well developed. In Latin there are no dual verb agreement forms, whereas in Homeric Greek there are (see e.g. Monro 1891).
28Note additionally that when the subject argument is elided, which it often is, the auxiliary and participle are inflected
CHAPTER 10. GRAMMATICAL PREDICTORS

(23) a. omnis=que noster equitatus eas cohortes
   est secutus. be.3.sg.prs.ind follow.ptcpl.prf.m.sg.nom
   ‘And all our cavalry followed those cohorts’ (Caes. Civ. 3. 68)

   b. plures eam sententiam secuti
      sunt be.3.pl.prs.ind
      ‘Many followed that opinion’ (Cic. Phil. 6. 3)

In infinitival personal PACS, the participle, but not the auxiliary, is targeted for number agreement relations. Accordingly it is lexical-headed. Thus in (24a), the singular pronominal subject eum is reflected in the -um ending of the participle, and in (24b) the plurality of the lexical subject iudices is encoded in the participial ending -os, but the inflectional form of the auxiliary remains the same for both numbers — esse.

(24) a. eum id esse secutum
   PRN.m.sg.acc PRN.n.sg.acc be.prs.inf follow.ptcpl.prf.m.sg.acc
   ‘That he followed it’ (Cic. Inv. 2. 8)

   b. id esse secutos iudices
   PRN.n.sg.acc be.prs.inf follow.ptcpl.prf.m.sg.acc judge.m.pl.acc
   ‘That the judges followed it’ (Cic. Clu. 97)

Clausal subjects and impersonal verbs are neutral with respect to number, and so take default (neuter) singular marking.

(25) a. Quo factum est ut plus quam
donde happen.ptcpl.prf.n.sg.nom be.3.sg.prs.ind that.c more than
colleagues Miltiades ualeret. colleague.m.pl.nom Miltiades.m.sg.nom prevail.3.sg.impf.sbjv

   b. *Quo facta sunt ut plus quam
donde happen.ptcpl.prf.n.pl.nom be.3.pl.prs.ind that.c more than
colleagues Miltiades ualeret. colleague.m.pl.nom Miltiades.m.sg.nom prevail.3.sg.impf.sbjv
   ‘By which it happened that Miltiades prevailed more than his colleague.’ (Nep. Milt. 5)

Impersonal verbs denoting multiple events are inflected as singular, with the event plurality marked by lexical means (such as bis in the example below). This is because the Latin verb is not inflected for verbal number (also termed pluractionality).

(26) a. bis cum iis pugnatum est
twice with prn.m.pl.abl fight.ptcpl.prf.n.sg.nom be.3.sg.prs.ind

   b. *bis cum iis pugnata sunt
twice with prn.m.pl.abl fight.ptcpl.prf.n.pl.nom be.3.pl.prs.ind
   ‘With them two fights took place’ (Liv. 38. 46. 3)

10.8.1.3.2 Grammatical effects There is, to my knowledge, next to nothing in the literature on the effect of grammatical number on internal AVC word order in the languages that display such variation. The only discussion that I can find is that of Steele (1913) concerning Latin PACS of the future infinitive type, e.g. facturum esse ~ esse facturum ‘to be going to do, make’, with reference to Livy (59 BC – AD 17), a historiographer writing in the Augustan period. He states, “Weissenborn, on Livy xxxix. 26. 12, calls attention to the fact that Livy very rarely places esse before the participle, citing xl. 41. 8 and for the grammatical number of the “understood” subject of the clause.
10.8. NUMBER

To these are to be added iii. 47. 7; xxxi. 13. 7; and xxxv. 35. 7 — all with the plural, for in the singular Livy steadfastly adheres to the order "-urum esse". Although these comments are unfortunately rather impressionistic and lacking in detail, one could say, roughly, that plural forms of the pac are associated with the variant est factus and singular forms are associated with factus est. In (27a) we have a plural future infinitive pac which is realized with an est factus variant, whilst in (27b) a singular form is realized with the factus est variant. These impressionistic comments clearly stand in need of a rigorous, quantitative assessment.

(27)  
(a. non spero esse passuros illos  
NEG hope.1.SG.PRS.IND be.PRS.IND allow.PTCPL.FUT.M.PL.ACC that.PRN.M.PL.ACC  
qui arma habent  
REL.M.PL.NOM arms.N.PL.ACC have.3.PRS.IND  
'I hope that those who have arms will not allow it' (Liv. 3. 47. 7)  
(b. omnia=que de sententia patrum facturum  
all.N.PL.ACC=and about feeling.F.SG.ABL father.M.PL.GEN do.PTCPL.FUT.M.SG.ACC esse  
be.PRS.IND  
'That he would do everything in accordance with the decision of the fathers' (Liv. 25. 7. 1)

10.8.2 Variable profile

In the foregoing we have seen that there is some cross-linguistic and cross-constructional support for supposing that the grammatical category of number might be relevant to the alternation at hand. It is therefore well justified from a theoretical and empirical perspective.

Following on from Steele’s (1913) comments, I annotate the dataset for the morphological number of the pac’s subject (whether overt or not). This is straightforwardly annotated, as it can be identified by examining the inflectional ending on the auxiliary and/or the participle.

We might thus hypothesize that singular pacs conspire to produce the factus est order, whereas plural pacs conspire to effect the est factus order. The variable is termed SubjectNumber.

10.8.3 Statistical modelling

Overall, there 1693 pacs inflected in the singular and 716 pacs inflected in the plural, a difference which in itself is highly significant ($\chi^2 = 396.23, df = 1, p < 0.001$), providing evidence that plural pacs are the marked category (at least in terms of frequency). Of course, there are differences between the different texts, as different texts are talking about different things: in some texts, plural entities may be as salient as singular entities, as in for instance the Bellum Hispaniense and Varro’s Res Rusticae, where there is a rate of plural pacs of 48.03% and 47.58% respectively. In Nepos, on the other hand, he is glorifying the deeds of individual people, hence why we might find such a strong asymmetry in this author’s works with only 17.93% being plural pacs.

Table 10.29 presents the relationship between grammatical number and serialization choice. The slight asymmetry discernable here, whereby singular forms are more likely to be realized as factus est, has some initial statistical support from a $\chi^2$-test of independence ($\chi^2 = 32.36, df = 1, p < 0.001$).

The four multilevel models to be compared with respect to their deviance and AIC are as follows. First, a varying intercepts model containing the predictor and varying intercepts for lemma and text is configured:

number.ri<-glmer(RealizationOfPAC~SubjectNumber+  
(1|LemmaParticiple)+(1|Text),pac,family=binomial)

29Weissenborn was a textual critic and commentator on Livy’s Ab Urbe Condita.
The second model is the same as the first, but with the addition of varying slopes for lemma and text.

\[
\text{number.rs.both} \leftarrow \text{glm} (\text{RealizationOfPAC} \sim \text{SubjectNumber} + \\
(\text{SubjectNumber} | \text{LemmaParticiple}) + (\text{SubjectNumber} | \text{Text}), \text{pac}, \text{family}=\text{binomial})
\]

The third model contains varying slopes for only lemma:

\[
\text{number.rs.lemma} \leftarrow \text{glm} (\text{RealizationOfPAC} \sim \text{SubjectNumber} + \\
(\text{SubjectNumber} | \text{LemmaParticiple}) + (1 | \text{Text}), \text{pac}, \text{family}=\text{binomial})
\]

And the fourth model contains varying slopes for only text:

\[
\text{number.rs.text} \leftarrow \text{glm} (\text{RealizationOfPAC} \sim \text{SubjectNumber} + \\
(1 | \text{LemmaParticiple}) + (\text{SubjectNumber} | \text{Text}), \text{pac}, \text{family}=\text{binomial})
\]

The comparisons of deviance and \textit{AIC} values, along with other pertinent \textit{anova} information, can be found in Table 10.30. This shows that the three more complex number models are significantly more informative than the model that contains only the number predictor and the varying intercepts (\text{number.ri} vs. \text{number.rs.lemma} (\chi^2 = 14.322, df_2, p=0.0007763; \text{number.ri} vs. \text{number.rs.text} (\chi^2 = 8.3583 df_2, p=0.01531; \text{number.ri} vs. \text{number.rs.both} (\chi^2 = 23.524 df_4, p=9.95e-05).

If we then compare the most complex model \text{number.rs.both} with the nested models \text{number.rs.lemma} and \text{number.rs.text}, the former is significantly more informative than the latter two, thus indicating that there is significant by-lemma and by-text variation in the slopes (respectively: \chi^2 = 9.202, df = 2, p = 0.01; \chi^2 = 15.165, df = 2, p < 0.001), with more variance for lemma (Var = 0.76, SD = 0.87) than for text (Var = 0.14, SD = 0.37). The additional complexity is therefore required, and it is confirmed by the fact that \text{number.rs.both} has the lowest \textit{AIC} of all the models reported in Table 10.30.

\[
\begin{array}{cccccc}
\text{Df} & \text{AIC} & \text{BIC} & \text{logLik} & \text{deviance} \\
\text{number.ri} & 4 & 3121.01 & 3144.16 & -1556.50 & 3113.01 \\
\text{number.rs.lemma} & 6 & 3110.69 & 3145.41 & -1549.34 & 3098.69 \\
\text{number.rs.text} & 6 & 3116.65 & 3151.37 & -1552.32 & 3104.65 \\
\text{number.rs.both} & 8 & 3105.48 & 3151.78 & -1544.74 & 3089.48 \\
\end{array}
\]

Table 10.30: ANOVA statistics for various \textit{glmer}s containing grammatical number information on the subject NP

This model is then compared with the baseline \textit{vcm}. As Table 10.31 demonstrates, the model with the predictor, the varying intercepts, and the varying slopes is significantly more informative than the variance components model.

\[
\begin{array}{ccccccc}
\text{Df} & \text{AIC} & \text{BIC} & \text{logLik} & \text{deviance} & \text{Chisq} & \text{Chi Df} & \text{Pr(>Chisq)} \\
\text{baseline.vcm} & 3 & 3139.41 & 3156.77 & -1566.70 & 3113.41 \\
\text{number.rs.both} & 8 & 3105.48 & 3151.78 & -1544.74 & 3089.48 & 43.92 & 5 & 0.0000 \\
\end{array}
\]

Table 10.31: ANOVA of baseline \textit{glm} and selected \textit{glmer} containing grammatical number information on the subject NP
The parameters of the selected model are reported in Table 10.32. This indicates that when the pac receives plural inflection as opposed to singular inflection, the logit probability of a factus est outcome decreases by $-0.49$, corresponding to an odds ratio of 0.61. The 95% confidence intervals of $[-0.83265088, -0.1452114]$ associated with this parameter estimate, while quite wide, are nonetheless in negative logit space, and the $p$-value computed off the absolute $z$-value of $-2.788$ and the relevant degrees of freedom indicates that it is highly significant at the $p < 0.01$ level. Grammatical number is therefore of substantive importance to the pac alternation, and in the direction claimed by Steele (1913).

|                | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 0.3354   | 0.1460     | 2.297   | 0.0216   |
| SubjectNumberplural | -0.4889 | 0.1754     | -2.788  | 0.0053   |

Table 10.32: Final glmer for serialization $\sim$ grammatical number

Let us now examine by-text and by-lemma variability in the slope estimates, as displayed graphically in Figure 10.3. The plot for individual by-text coefficients show that all of the texts bar one (Cicero’s Phillipics) are firmly in negative logit space (decreasing the likelihood of a factus est outcome). Concerning verb lemma, while most of the verb lemmas are in negative logit space, there is more variability. Specifically, 17/604 lemmas exhibit positive coefficients, most strikingly for mitto ‘send’. For these verb lemmas, this means that if the pac is inflected in the plural, it actually increases the likelihood of a factus est outcome.

10.9 Gender

We now move on to a second morphological/grammatical category, namely that of gender.

10.9.1 Linguistic importance of gender

10.9.1.1 General background

Like number, gender is also important to the unmarked vs. marked distinction in language. Here, the masculine is usually considered the unmarked gender (Greenberg, 1966: 39), though it has been noted that this may not be entirely accurate crosslinguistically (Croft, 2003: 100). By contrast, the neuter gender is regarded as the most marked. Thus, if a language has the neuter gender it also has the feminine gender, and if a language has the feminine gender, it also has the masculine gender. We have the following markedness hierarchy: neuter > feminine > masculine. This markedness distinction has a number of linguistic consequences.

Greenberg (1966: 31) notes that in conjoined noun phrases comprising e.g. a masculine conjunct and a feminine conjunct the modifying element, where there is one, is typically in the unmarked gender, i.e. masculine. The example given is that of Spanish el hijo y la hija son buenos ‘The son and the daughter are good’ — here buenos ‘good’ is inflected in the masculine. In Semitic languages, the unmarked masculine form is zero realized whereas the marked feminine gender is realized by a suffix (Greenberg, 1966: 39). In Latin, in the first declension (largely feminine), the distinction between the nominative and vocative singular is neutralized — for instance, the nominative and vocative form of ‘girl’ is puella. By contrast, in the second declension masculine singular, this distinction is marked, with for instance seru-us ‘slave’ indicating the nominative and seru-e indicating the vocative. There is further neutralization in second declension neuter forms in both the singular and the plural, where nominative, vocative, and accusative forms are marked with -um and -a respectively. Finally, Greenberg (1966) presents frequency information from Spanish adjectives. Here, 62.7% are masculine while 37.3% are feminine, pointing to the strong asymmetry in gender marking in this language.
Figure 10.3: Individual by-text (left) and by-lemma (right) coefficients for GLMER containing grammatical number information
10.9. GENDER

10.9.1.2 Effects of gender on variable grammatical realization

To my knowledge there is nothing in the literature on the effect of gender on variable grammatical realization. Nonetheless, it cannot be doubted that it is an important grammatical feature across languages of the world (Croft, 2003).

10.9.1.3 Grammatical gender in AVCS and in Latin PACS

10.9.1.3.1 Morphological encoding Grammatical gender is also relevant to the formal structure of AVCS, and much like number can be headed in different loci of the construction (Corbett, 2000). In Latin, a three-way gender distinction is made between masculine, feminine, and neuter, and this information is expressed in the inflectional ending of the word. For instance, with first declension nouns in the nominative singular, we find respectively, e.g. *seru-us* ‘slave (masc.)’, *seru-a* ‘slave woman (fem.)’, and *templ-um* ‘temple (neut.)’. For PACS, the grammatical gender of the head of the subject NP spreads to the verbal predicate via agreement rules, just as with number. However, unlike number, gender information is located only on the participle. Adopting Corbett’s (2000) terminology it is thus lexically headed. We illustrate this in (28a)–(28c). Note that the form of the auxiliary does not depend on the grammatical gender of the subject — in these examples it is *est* throughout.

(28) a. [Antonius] ... edem die quo [0]
   Antonius.M.SG.NOM ... same.M.SG.ABL day.M.SG.ABL REL.M.SG.ABL [0]
   profectus est reuertitur ← masculine
depart.PTCPL.PRF.M.SG.NOM be.3.SG.PRS.IND return.3.SG.PRS.IND
   ‘Antonius returned on the same day on which he departed’ (Caes. Civ. 1. 18. 3)

b. his ex manubiis [arx Athenarum] ...
   this.PRN.F.PL.ABL from prize.funds.PRN.F.PL.ABL citadel.F.SG.NOM Athens.F.PL.GEN ...
   est ornata ← feminine
   be.3.SG.PRS.IND adorn.PTCPL.PRF.F.SG.NOM
   ‘From these prize funds, the citadel of Athens was adorned’ (Nep. Cim. 2. 5)

c. itaque [aes alienum provinciae] eo
   therefore bronze.N.SG.NOM foreign.N.SG.NOM province.F.SG.GEN PRN.N.SG.ABL
   biennio multiplicatum est. ← neuter
   two.year.period.N.SG.ABL increase.PTCPL.PRF.N.SG.NOM be.3.SG.PRS.IND
   ‘The debt of the province increased in that two year period’ (Caes. Civ. 3. 32. 5)

Impersonal PACS, as in (29a), and clausal CP subjects, as in (29b), receive neuter marking.

(29) a. reliquis oppidi partibus sic est
   remaining.F.PL.ABL town.N.SG.GEN part.F.PL.ABL thus be.3.SG.PRS.IND
   pugnatum ut
   fight.PTCPL.PRF.N.SG.NOM that.C
   ‘In the remaining parts of the town fighting took place to such a degree that …’ (Caes. Civ. 3. 112. 7)

b. quo factum est [ut Atheniensium
   wherefore happen.PTCPL.PRF.N.SG.NOM be.3.SG.PRS.IND that.C Athenian.M.PL.GEN
   muri ex sacellis sepulcris=que constarent.]
   wall.M.PL.NOM from shrine.N.PL.ABL tomb.N.PL.ABL=and consist.of.3.PL.IMPF.SBJV
   ‘As a result it happened that the walls of the Athenians consisted of shrines and tombs’ (Nep. Them. 6. 5)

30In this example the subject is elided, denoted as [0], and is coreferential with the matrix clause subject *Antonius*. This example has been changed slightly from the original text.
10.9.2 Variable profile and operationalization

Although gender is less well motivated theoretically and empirically compared to number (and person, see below), a predictor variable is nonetheless configured for the grammatical gender of the PAC’s subject. The information contained in the participle’s ending is used to code for three levels masculine, feminine and neuter. As displayed in Table 10.33, suffixes in -i and -os are explicitly masculine and suffixes in -am, -ae and -as are explicitly feminine, and were automatically tagged; by contrast, suffixes in -um are — without syntactic and contextual clues — ambiguous between masculine (accusative singular) and neuter (nomina
tive singular or accusative singular) and suffixes in -a are — again without syntactic and contextual clues — ambiguous between feminine (nomina
tive singular) and neuter (nomina
tive plural or accusative plural), and were thus hand-coded. This variable is termed SubjectGender. For examples, I refer the reader to (28) above.

10.9.3 Statistical modelling

The overall distribution of gender’s effect on serialization choice is given in Table 10.34. There appears to be a slight effect whereby PACs inflected in the feminine gender are dispreferred with factus est serializations compared with masculine and neuter genders, but a χ²-test of independence does not support this association (χ² = 5.6843, df = 2, p = 0.0583). We will see below that simply using basic statistical tests for crosstabulation that do not take into account the multilevel structure of the present data sample can result in Type II (conservative) statistical errors, which suggest there is no effect when in fact there is.

<table>
<thead>
<tr>
<th>Case</th>
<th>Masculine</th>
<th>Feminine</th>
<th>Neuter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>Nominative</td>
<td>-us</td>
<td>-a</td>
</tr>
<tr>
<td></td>
<td>Accusative</td>
<td>-um</td>
<td>-a</td>
</tr>
<tr>
<td>Plural</td>
<td>Nominative</td>
<td>-i</td>
<td>-ae</td>
</tr>
<tr>
<td></td>
<td>Accusative</td>
<td>-os</td>
<td>-as</td>
</tr>
</tbody>
</table>

Table 10.33: Gender endings of the participle in the PAC

<table>
<thead>
<tr>
<th>Serialization</th>
<th>factus est (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>masculine</td>
<td>487 (38.87)</td>
<td>766 (61.13)</td>
<td>1253</td>
</tr>
<tr>
<td>feminine</td>
<td>208 (44.64)</td>
<td>258 (55.36)</td>
<td>466</td>
</tr>
<tr>
<td>neuter</td>
<td>264 (38.26)</td>
<td>426 (61.74)</td>
<td>690</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
<td>2409</td>
</tr>
</tbody>
</table>

Table 10.34: Relationship between grammatical gender and PAC serialization

The four multilevel models to be compared with respect to their deviance and AIC are as follows. First, a varying intercepts model containing the gender predictor and varying intercepts for lemma and text is configured:

```
gender.ri<-glmer(RealizationOfPAC~SubjectGender+(1|LemmaParticiple)+(1|Text),pac,family=binomial)
```

The second model is the same as the first, with additional terms for varying slopes for lemma and text.

```
gender.rs.both<-glmer(RealizationOfPAC~SubjectGender+(SubjectGender|LemmaParticiple)+(SubjectGender|Text),pac,family=binomial)
```

The third model contains varying slopes for only lemma:

```
gender.rs.lemma<-glmer(RealizationOfPAC~SubjectGender+(SubjectGender|LemmaParticiple)+(1|Text),pac,family=binomial)
```

31 They can also be anti-conservative.
And the fourth model contains varying slopes for only text:

\[ \text{gender.rs.text} <- \text{glmer(RealizationOfPAC~SubjectGender} + (1|LemmaParticiple)} + (\text{SubjectGender}|\text{Text}), \text{pac,family=binomial}) \]

The model comparisons are instructive (cf. Table 10.35). The varying intercepts only model \text{gender.ri} fares significantly poorer than \text{gender.rs.lemma} \((\chi^2 = 16.924, df = 5, p = 0.004646)\) and \text{gender.rs.both} \((\chi^2 = 20.14, df = 10, p = 0.02795)\), but not \text{gender.rs.text} \((\chi^2 = 4.2219, df = 5, df = 0.5179)\), suggesting significant differences for by-lemma slopes but not for by-text slopes. This is confirmed by comparing the fullest model \text{gender.rs.both} with the two nested varying intercept and varying slope models. To illustrate, \text{gender.rs.both} exhibits a significant deviance reduction of \(3125.2 - 3109.3 = 15.918\) \((df = 5, p = 0.007081)\) from \text{gender.rs.text}, but the deviance reduction of \(3112.5 - 3109.3 = 3.2162\) between it and \text{gender.rs.lemma} is non-significant \((df = 5, p = 0.6667)\). In sum, the addition of varying by-text slopes adds nothing of substantive interest. Model \text{gender.rs.lemma}, being the most informative and parsimonious at the same time, is then inspected for its significance against the baseline vcm (cf. Table 10.36). This indicates that the model that includes the predictor significantly reduces the deviance as compared with the variance components model that includes only lemmatic and textual information.

<table>
<thead>
<tr>
<th>Model</th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender.ri</td>
<td>5</td>
<td>3139.46</td>
<td>3168.40</td>
<td>-1564.73</td>
<td>3129.46</td>
</tr>
<tr>
<td>gender.rs.lemma</td>
<td>10</td>
<td>3132.54</td>
<td>3190.41</td>
<td>-1556.27</td>
<td>3112.54</td>
</tr>
<tr>
<td>gender.rs.text</td>
<td>10</td>
<td>3145.24</td>
<td>3203.11</td>
<td>-1562.62</td>
<td>3125.24</td>
</tr>
<tr>
<td>gender.rs.both</td>
<td>15</td>
<td>3139.32</td>
<td>3226.12</td>
<td>-1554.66</td>
<td>3109.32</td>
</tr>
</tbody>
</table>

Table 10.35: ANOVA statistics for various glmers containing grammatical gender information on the subject np

<table>
<thead>
<tr>
<th>Model</th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.vcm</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gender.rs.lemma</td>
<td>10</td>
<td>3132.54</td>
<td>3190.41</td>
<td>-1556.27</td>
<td>3112.54</td>
<td>20.87</td>
<td>7</td>
<td>0.0040</td>
</tr>
</tbody>
</table>

Table 10.36: ANOVA of baseline glm and selected glmer containing grammatical gender information on the subject np

With the model now established, we are now able to interpret the fixed-effect parameters for the predictor variable, which are displayed in Table 10.37. As usual, the logit probability is in terms of a factus est outcome, and the input is configured such that masculine is the baseline level to which feminine and neuter are to be compared, given that it is the factor level with the most number of observations sampled on it \((n = 1253)\). The parameter for feminine PACs indicates that when the PAC is inflected for feminine gender as opposed to masculine gender the logit probability of a factus est outcome decreases by \(-0.24155\), equivalent to an odds ratio of 0.7854095. The coefficient divided by its standard error gives a \(z\)-value of \(-2.041\). Since its absolute value is above the critical \(z\)-value of 1.96 for large sample sizes, we can conclude with 95.88% confidence that the parameter estimated from the sample is reflective of that for the population. The confidence intervals associated with this estimate are also indicative of the robustness of the effect, as they lie completely in negative logit space (from \(-0.47\) to \(-0.01\)), although it is to be noted that they are quite wide. In other words, using a feminine PAC makes it less likely to obtain a factus est serialization than when using a masculine PAC. By contrast, no such effect is found for the parameter concerning neuter gender: they appear to behave identically to masculine PACs. This can be seen from the fact that the 95% confidence intervals cross zero in logit space (from \(-0.3104159\) to 0.204336833), and the absolute \(z\)-value is well below 1.96.

The model that has been selected suggests that there is significant by-lemma variability with respect to the slopes. These are plotted in Figure 10.4. For the level feminine vs. masculine, the cloud is mostly in
negative logit space, promoting est factus. There are 11 lemmas which show a polar opposite tendency to the majority: coepi ‘begin’, cognosco ‘find out’, facio ‘do, make’, sum ‘be’, contineo ‘contain’, demonstro ‘point out’, venio ‘come’, pugno ‘fight’, insero ‘graft’ and sero ‘sow’. If the PAC is one of these verb lemmas and is inflected for feminine, a factus est outcome is stimulated.

By contrast, the cloud of verb lemmas for neuter is not so clearly situated in either negative or positive logit space, with 255 above zero and 349 below zero. There is far too much variability for neuter slopes for anything general to be said about their effect. These plots illustrate effectively why feminines are important, but not neuters.

10.9.4 Discussion

To my knowledge this is the first study that has assessed the relevance of gender on variable grammatical choices. But why exactly should it have an effect? One reason that Benedikt Szmrecsanyi (p.c.) suggests to me is that, given that masculines are more of a “default” gender, they may be easier to access. Hence, if the participle is marked for masculine it comes earlier in the sequence more readily.
10.10. Person

10.10.1 Introduction

Person is a third morphological contrast that is of linguistic consequence (see Siewierska (2004) for a summary). It is typical to find a distinction between local (first and second) persons and non-local (third) persons, but richer systems are to be found throughout the world’s languages. Here we concentrate on the effect grammatical person might have on the Latin PAC alternation.

10.10.2 Linguistic importance of person

10.10.2.1 Effects of person on variable grammatical realization

Several categorical effects of person are summarized by Aissen (1999: with references). For example, Nocte (Tibeto-Burman family) makes a distinction between direct and inverse verbs. In this language, when the subject is a first or second person form and the object is third, the verb is marked as direct; when the subject is third person and the object is first or second, the verb is marked as indirect.

Probabilistic effects of grammatical person are also found. Recent work by Alcorn (2009), on the word order difference [preposition + personal pronominal object] vs. [personal pronominal object + preposition] in Old English, finds a strong effect of grammatical person, a point she reports is first mentioned by Wende (1915: 76). In third person forms the pronoun is statistically significantly placed to the left of the preposition in a [personal pronominal object + preposition order] configuration, while this position is strongly dispreferred when it is non-third-person (i.e. first and second person). This asymmetry is shown to be independent of other possible influences on the placement of the pronoun with respect to its preposition, such as, i.a., animacy, narrative mode, case, and the lexis and semantics of the preposition. This finding is expanded on and confirmed in her PhD thesis (Alcorn, 2011: 233–240; cf. 79f. for the operationalization and monovariate results), when other variables are controlled for in a logistic regression (using GoldVarb).

10.10.2.2 Grammatical number in AVCs and in Latin PACs

10.10.2.2.1 Morphological encoding

Like number and gender, person information is also a featural property of AVCs, as Anderson (2006) mentions. In Latin a three-way distinction is made between first, second, and third person,32 preserved in pronominal forms (e.g. ego, me, mei, mihi, me, nos, nostri/nostrum, nobis ‘I, me, etc.’ vs. tu, te, tui, tibi, te, uos, uestri/uestrum, uobis ‘you’ vs. se/sese, suus, sibi, se/sese ‘him/her/it/them’, amongst others such as the demonstratives is, hic, iste and ille). Person information is transmitted to the verbal predicate via agreement rules, and as such the PAC is inflected for this morphological category. In PACs, only the finite auxiliary is inflected for person information, with the participle unmarked. Consequently, we might say that it is auxiliary-headed in Anderson’s (2006) terminology. Controlling for other featural characteristics of PAC inflection, we witness the three-way distinction in the following examples, with first person exemplified in (30a), the second person in (30b), and the third person in (30c).

(30) a. ipse in Asian profectus sum
    self.f.sg.nom into Asia.f.sg.acc depart.ptcpl.prf.m.sg.nom be.1sg.prs.ind
    ‘I departed for Asia’ (Cic. Att. 5. 21. 7)

b. tu ... Roma profectus es
    prn.2.m.sg.nom ... Rome.f.sg.abl depart.ptcpl.prf.m.sg.nom be.2sg.prs.ind
    ‘You departed from Rome’ (Cic. Att. 3. 9. 3)

32 Though third persons can be split into proximal forms, such as hic ‘this near me’, iste ‘that near you’, and distal forms ille ‘that over there’.
c. ille Idibus a me profectus
that.PR.N.M.SG.NOM Ides.F.PL.ABL from PRN.1.SG.ABL depart.PTCPL.PRF.M.SG.NOM
est be.3SG.PRS.IND
‘That man departed from me on the Ides’ (Cic. Att. 9. 9. 3)

By contrast, in nonfinite PACs, person is encoded neither on the auxiliary nor on the participle. We might term it zero-headed in the spirit of Anderson’s (2006) terminology, as the following contrasting examples demonstrate:

(31) a. me in Asiam profectus esse
    (Cic. Att. 5. 21. 7)
b. te . . . Roma profectus esse
    (Cic. Att. 3. 9. 3)c. eum Idibus a me profectus esse
    (Cic. Att. 9. 9. 3)

10.10.2.2 Grammatical effects  To my knowledge nothing has been reported in the literature on the effect of person on serialization in AVCs or PACs more specifically.

10.10.3 Variable profile and operationalization

Predictor information is annotated for the person value of the subject NP. This information is annotated directly from the form of the auxiliary in finite forms; in nonfinite forms manual annotation was required, because they are not marked (see above). We initially identify three levels first, second, third, and term the variable SubjectPerson. (For examples, see above.) However, first (77/2409, i.e. 3.20%) and second persons (55/2409, i.e 2.28%) are quite infrequent, probably because of the nature of the texts, so it was decided to collapse these to create a binary factor with the levels local and nonlocal.

10.10.4 Statistical modelling

The overall relationship between this grammatical person predictor and serialization choice is presented in Table 10.38. There is a slight asymmetry, with local (first and second) person subjects associating slightly more frequently with factus est serializations, but a $\chi^2$-test of independence does not support such an association ($\chi^2 = 1.66$, $df = 1$, $p = 0.197$).

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>nonlocal</td>
<td>914 (40.14)</td>
<td>1363 (59.86)</td>
<td>2277</td>
</tr>
<tr>
<td>local</td>
<td>45 (34.09)</td>
<td>87 (65.91)</td>
<td>132</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
<td>2409</td>
</tr>
</tbody>
</table>

Table 10.38: Relationship between grammatical person and PAC serialization

The four multilevel models to be compared with respect to their deviance and $AIC$ are as follows. First, a varying intercepts model containing the predictor and varying intercepts for lemma and text is configured:

```r
person.ri<-glmer(RealizationOfPAC~SubjectPerson+
                   (1|LemmaParticiple)+(1|Text),pac,family=binomial)
```

The second model is the same as the first, but with the addition of varying slopes for lemma and text.

---

33In Cicero’s *Philippics* where Cicero (first person) opposes himself to Antonius (second person), there is naturally a higher rate of first and second persons.
Morphosyntactic category of the preverbal expression

10.11.1 Introduction

This section motivates the inclusion in the information space of a predictor concerning the grammatical category of the last preverbal expression. I begin in Section 10.11.2 by familiarizing the reader with what has been said on this subject in the general verb cluster literature that adopts quantitative techniques (Section 10.11.2.1) and for the Latin pac in (Section 10.11.2.2). I discuss the operationalization of the relevant variables in Section 10.11.3, report the statistical model in Section 10.11.4, and offer a qualitative discussion in Section 10.11.5.

10.11.2 Background

10.11.2.1 AVC literature

In his study on Dutch verb clusters containing a participle and an auxiliary, De Sutter (2009) examines the effect of the “information value” of the last preverbal word. Grammatical categories are grouped according
to how much referential content they contribute to the sentence: nouns, verb, and numerals are classed as being “highly informational”; adjectives and adverbs are classed as “intermediately informational”; conjunctions, prepositions, and pronouns are “low informational”. In a standard (i.e. all fixed-effects) GLM, he observes that when the last preverbal word belongs to a grammatical category in the highly informational group the odds of a participle-final outcome are 1.94 times greater than when the last word before the verb cluster belongs to a grammatical category from the low informational group (2009: 245).

Another study that has investigated the effect of the grammatical category of the last preverbal expression is Sapp (2011) study on historical German verb clusters, which vary between finite-dependent ‘1-2’ and dependent-finite ‘2-1’ orders. Drawing on observations from previous research, he assesses whether the full range of parts of speech that occur before the verb cluster have any influence on the word order inside the complex. Restricting this review to his Middle High German dataset (2011: 21–23), he finds that preverbal nouns, quantified noun phrases and adjectives prefer an association with the 1-2 order. By contrast, pronouns are associated with dependent-finite orderings. The explanation for this, he submits following the literature, is possibly due to phonological contouring triggered by the preverbal word: a phonologically light word preceding the verb cluster triggers a phonologically heavier word in the verb cluster, i.e. the dependent verb, to appear after it; by contrast, a phonologically heavy word preceding the verb cluster triggers a phonologically lighter word in the verb cluster, i.e. the finite verb, to follow it. When he recodes some of the categories as unstressed (pronoun, short adverbs, etc.) and others as stressed, he finds a significant effect, which may point to the fact that grammatical category is epiphenomenal to phonological constraints.

Thus in cross-language literature various grammatical aspects of the last preverbal expression exert influences on serialization choice within AVCS, though in some cases they may be epiphenomenal to other factors. Let us now turn to what has been said for Latin.

10.11.2.2 PAC literature

There has been relatively little discussion on the effect of preverbal constituency in the PAC literature. However, researchers on the grammar of Latin have observed that certain types of grammatical category regularly “attract” the verb esse, whether in copula or auxiliary function, such that they are adjacent to each other. One prominently mentioned category in this regard is the negative adverb non (Marouzeau, 1910; Lease, 1919; Kroll, 1921; Müller, 1924; Hofmann et al., 1965; Adams, 1994; Devine & Stephens, 2006). Adams (1994: 9) remarks as follows:

\[
\text{[E]}\text{sse is regularly attracted to one particular host, namely the negative non and some other negatives. The nexus between non and esse is shown by the fact that a negated passive perfectum, gerundive, periphrastic future or adjectival predicate regularly takes the form (e.g.) factum non est or non est factum rather than non factum est.}
\]

Next to the negative, there are some further grammatical categories which have been claimed to attract esse. These are temporal adverbs (Adams, 1994: 38f.), demonstratives (Adams, 1994: 25–28, 37.), modifiers involved in comparison (Adams, 1994: 24, 36), quantifiers, numerals, and measuring expressions (Adams, 1994: 19f, Devine & Stephens, 2006: 182, 191), relative pronouns (Adams, 1994: 44-53, Devine & Stephens, 2006: 191), degree expressions (Adams, 1994: 28, 38, Devine & Stephens, 2006: 182, 191), complementizers (Vogel, 1937, Devine & Stephens, 2006), and interrogatives (Devine & Stephens, 2006: 191). If the preverbal expression includes one or more of these esse-attracting categories, then it follows that the PAC should be realized as est factus instead of factus est.

The assumption why the negative and the other grammatical categories attract esse is that they are typically “focused elements” (Adams, 1994: 34). As we will see in another chapter, the information structural notion of “focus” is another and possibly more explanatory factor also mentioned in the PAC literature, so the influence of these grammatical categories on the choice on construction could well be epiphenomenal to information structural constraints.
Finally, if nothing precedes the PAC, factus est is demonstrably preferred over est factus (Vogel, 1937: 35, 45, 50, 55).

10.11.3 Variable profile and operationalization

Rather than annotate a variable for the last preverbal word, as might be suggested by the previous PAC literature, I will instead annotate for the last preverbal category more generally, whether it is a word/larger constituent/clause (PreverbalCategory). The following levels were established: connectives and conjunctions (connective, n = 69), complementizers (C, n = 135), negative particle non ‘not’ and numquam ‘never’ (negative, n = 51), temporal adverbial phrases (temporalAdvP, n = 49), degree adverbial phrases (degreeAdvP, n = 84), other types of adverbial phrase (AdvP, n = 120), bare nouns whether as an NP or PP (noun, n = 447), bare demonstratives whether as an NP or PP (demonstrative, n = 82), bare quantifiers whether as an NP or PP (quantifier, n = 40), bare relatives whether as an NP or PP (relative, n = 132), bare personal pronouns whether as an NP or PP (personalPronoun, n = 60) noun phrase (NP, n = 538), adjective phrase (AdjP, n = 40), prepositional phrases (PP, n = 230), clause (clause, n = 212), participle phrases (PartP, n = 50), and nothing (nothing, n = 70)

10.11.4 Statistical modelling

The overall relationship between preverbal grammatical category and serialization choice is presented in Table 10.41. There are some strong skews in the data here, with, for instance, preverbal connectives and empty preverbal slots strongly selecting factus est serializations at a rate of 86.96% and 91%, respectively, and negatives strongly preferring est factus serializations at a rate of 82.35%. The asymmetry in the distribution is obviously significant on a χ²-test of independence (χ² = 176.5006, df = 16, p < 0.001)

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>204 (37.92)</td>
<td>334 (62.08)</td>
<td>538</td>
</tr>
<tr>
<td>AdjP</td>
<td>22 (55.00)</td>
<td>18 (45.00)</td>
<td>40</td>
</tr>
<tr>
<td>clause</td>
<td>43 (20.28)</td>
<td>169 (79.72)</td>
<td>212</td>
</tr>
<tr>
<td>connective</td>
<td>9 (13.04)</td>
<td>60 (86.96)</td>
<td>69</td>
</tr>
<tr>
<td>AdvP</td>
<td>46 (38.33)</td>
<td>74 (61.67)</td>
<td>120</td>
</tr>
<tr>
<td>negative</td>
<td>42 (82.35)</td>
<td>9 (17.65)</td>
<td>51</td>
</tr>
<tr>
<td>temporalAdvP</td>
<td>15 (30.61)</td>
<td>34 (69.39)</td>
<td>49</td>
</tr>
<tr>
<td>demonstrative</td>
<td>39 (47.56)</td>
<td>43 (52.44)</td>
<td>82</td>
</tr>
<tr>
<td>noun</td>
<td>202 (45.19)</td>
<td>245 (54.81)</td>
<td>447</td>
</tr>
<tr>
<td>quantifier</td>
<td>21 (52.50)</td>
<td>19 (47.50)</td>
<td>40</td>
</tr>
<tr>
<td>relative</td>
<td>62 (46.97)</td>
<td>70 (53.03)</td>
<td>132</td>
</tr>
<tr>
<td>C</td>
<td>78 (57.78)</td>
<td>57 (42.22)</td>
<td>135</td>
</tr>
<tr>
<td>PP</td>
<td>93 (40.43)</td>
<td>137 (59.57)</td>
<td>230</td>
</tr>
<tr>
<td>degreeAdvP</td>
<td>47 (55.95)</td>
<td>37 (44.05)</td>
<td>84</td>
</tr>
<tr>
<td>personalPronoun</td>
<td>19 (31.67)</td>
<td>41 (68.33)</td>
<td>60</td>
</tr>
<tr>
<td>nothing</td>
<td>6 (8.57)</td>
<td>64 (91.43)</td>
<td>70</td>
</tr>
<tr>
<td>PartP</td>
<td>11 (22.00)</td>
<td>39 (78.00)</td>
<td>50</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
<td>2409</td>
</tr>
</tbody>
</table>

Table 10.41: Relationship between preverbal grammatical category and PAC serialization

For this predictor, a single multilevel model was set up, because of the many number of factor levels resulting in model convergence problems when random slopes are included. Only random intercepts for the verb’s lemma and for the text are therefore included.

preverbal.ri<-glmer(RealizationOfPAC~PreverbalCategory+ (1|LemmaParticiple)+(1|Text),pac,family=binomial)
An ANOVA comparing this model with the baseline vcm is reported in Table 10.42. As this ANOVA clearly shows, the addition of the predictor significantly reduces the deviance from 3139.41 to 3002.63 (a reduction of 168.78), and the model preverbal.ri has the lower AIC. We should therefore accept it for an inspection of the parameters.

|        | DI  | AIC  | BIC  | logLik | deviance | Chisq | Chi DF | Pr(>|Chisq|) |
|--------|-----|------|------|--------|----------|-------|--------|----------|
| baseline.vcm | 3   | 3139.41 | 3156.77 | -1566.70 | 3133.41 |       |        |          |
| preverbal.ri  | 19  | 3002.63 | 3112.58 | -1482.32 | 2964.63 | 16    | 0.0000 |          |

Table 10.42: ANOVA of baseline GLM and selected GLMER containing information on the grammatical category of the preverbal expression

The model with its fixed effects is reported in Table 10.43. The logit probability estimate is in terms of a factus est outcome. The predictor variable is coded with NPs as the baseline level (= 0), as it has the greatest number of observations recorded on it, against which the remaining levels are contrasted via treatment coding. The intercept indicates that NPs have an logit probability of 0.27245 (in odds 1.31), meaning that they occur 56.76% of the time with factus est. There are several completely irrelevant predictor levels: other adverbial phrases, temporal adverbial phraes, and prepositional phrases, and bare personal pronouns. These do not significantly differ from NPs with respect to the probability of factus est. There is a second batch of factor levels which are marginally significant at \( p < 0.1 \): specifically, these are bare demonstratives, nouns, quantifiers, and relatives. Given more data, these factor levels may be brought into statistically significant territory.

Eight parameter estimates are significantly different from the baseline level of NPs, with clauses, connectives, nothing, and participle phrases exhibiting positive estimates, thereby indicating a stronger preference for factus est, and negatives, complementizers, and degree adverbial phrases exhibiting negative coefficients, thereby indicating a stronger preference for est factus. Let us now run through these coefficients in terms of decreasing logits. First, when there is no preverbal expression as opposed to when it is an NP, the logit probability of a factus est outcome increases by 1.75471 (an odds ratio of 11.27), and the confidence intervals associated with this estimate lie between 0.8759335 and 2.63349643 on the logit scale. This is in the direction claimed by the previous literature. Second, when the preverbal expression is a connective/coordinator as opposed to when it is an NP, the logit probability of a factus est outcome increases by 1.34963 (an odds ratio of 11.27), and the confidence intervals associated with this estimate lie between 0.6052095 and 2.09405866 on the logit scale. Third, when the preverbal expression is a participle phrase as opposed to when it is an NP, the logit probability of a factus est outcome increases by 0.84113 (an odds ratio of 11.27), and the confidence intervals associated with this estimate lie between 0.1279343 and 1.55432012 on the logit scale. Fourth, when the preverbal expression is a clause as opposed to when it is an NP, the logit probability of a factus est outcome increases by 0.79396 (an odds ratio of 11.27), and the confidence intervals associated with this estimate lie between 0.3951288 and 1.19279464 on the logit scale. Fifth, when the preverbal expression is a complementizer as opposed to when it is an NP, the logit probability of a factus est outcome decreases by −0.69081 (an odds ratio of 11.27), and the confidence intervals associated with this estimate lie between -1.0969463 and -0.28467175 on the logit scale. This is in the direction claimed by the previous literature, as degree phrases have been claimed to attract esse towards them. Sixth, when the preverbal expression is an adverbial phrase of degree as opposed to when it is an NP, the logit probability of a factus est outcome decreases by −0.86770 (an odds ratio of 11.27), and the confidence intervals associated with this estimate lie between -1.3526539 and -0.38275240 on the logit scale. This is in the direction claimed by the previous literature, as degree phrases have been claimed to attract esse towards them. Seventh, when the preverbal expression is an adjective phrase as opposed to when it is an NP, the logit probability of a factus est outcome decreases by −0.87778 (an odds ratio of 11.27), and the confidence intervals associated with this estimate lie between -1.5561361 and -0.19942030 on the logit scale. Finally, when the preverbal expression is a negative adverb as opposed to when it is an NP, the logit probability of a factus est outcome decreases by
−2.30515 (an odds ratio of 11.27), and the confidence intervals associated with this estimate lie between -3.0798125 and -1.53048953 on the logit scale. This is in the direction claimed by the previous literature. In sum, there are robust and reliable effects of the preverbal category on the choice on PAC serialization, when lemmatic and textual differences are taken into account.

|                          | Estimate | Std. Error | z value | Pr(>|z|) |
|--------------------------|----------|------------|---------|----------|
| (Intercept)              | 0.27245  | 0.17167    | 1.587   | 0.112497 |
| PreverbalCategoryAdjP    | -0.87778 | 0.34611    | -2.536  | 0.011208 |
| PreverbalCategoryclause | 0.79396  | 0.20349    | 3.902   | 9.55e-05 |
| PreverbalCategoryconnective | 1.34963 | 0.37982    | 3.553   | 0.000380 |
| PreverbalCategoryAdvP    | -0.01461 | 0.21890    | -0.067  | 0.946786 |
| PreverbalCategorynegative | -2.30515 | 0.39524    | -5.832  | 5.47e-09 |
| PreverbalCategorytemporalAdvP | 0.30596 | 0.33688    | 0.908   | 0.363763 |
| PreverbalCategorydemonstrative | -0.43273 | 0.24843    | -1.742  | 0.081532 |
| PreverbalCategorynoun    | -0.25442 | 0.13631    | -1.867  | 0.061971 |
| PreverbalCategoryquantifier | -0.58610 | 0.34506    | -1.699  | 0.094030 |
| PreverbalCategoryrelative | -0.36450 | 0.20506    | -1.778  | 0.075482 |
| PreverbalCategoryC       | -0.69081 | 0.20722    | -3.334  | 0.000857 |
| PreverbalCategoryPP      | -0.14001 | 0.16844    | -0.831  | 0.405853 |
| PreverbalCategorypersonalPronoun | 0.13477 | 0.30522    | 0.442   | 0.658812 |
| PreverbalCategorynothing | 1.75471  | 0.44837    | 3.914   | 9.09e-05 |
| PreverbalCategoryPartP   | 0.84113  | 0.36388    | 2.312   | 0.020803 |

Table 10.43: Final glmer for serialization ∼ preverbal grammatical category

10.11.5 Discussion

The effects established here do not tally precisely with the claims of the previous literature. Negatives, complementizers, degree adverbial phrases, and empty preverbal slots are well-behaved in this respect, but temporal adverbial phrases are not (in fact the actual sample coefficient indicates higher likelihood for *factus est*, which is the reverse of the claim, but it is not statistically significant). As for bare demonstratives, quantifiers, and relatives, it is interesting that they do not reach the standard threshold of statistical significance of < 0.05, given the importance the previous literature has attached to such grammatical categories. However, they may reach significance if the sample size is increased, which points to the fairly weak effect of these levels. It has also been demonstrated here that several levels — which go unmentioned in previous studies — are of importance: specifically, participle phrases, clauses, and connectives strongly associate with *factus est*, whereas adjectival phrases associate more with *est factus*.

10.12 Clause type

10.12.1 Introduction

Clause type is a mainstream syntactic determinant of grammatical variation Siewierska (1988: 89ff.). The importance of such factors is witnessed in their categorical application in the syntax of some languages. For instance, in counterfactual conditional clauses in English, the auxiliary *had* goes in initial position when the complementizer is absent, whereas when the complementizer is realized then the auxiliary is realized in its typical post-subject position. Similarly, in German and some other West Germanic languages, in main clauses finite verbs occur in second position, while in subordinate clauses the verb is realized in final position. However, in other languages, such factors operate probabilistically. Against this backdrop, we will first examine what researchers into AVC/PAC variability have had to say about clause type effects (in Section 10.12.2); then in Section 10.12.3 I present the operationalization of this
variable; Section 10.12.4 reports the statistical model, and in Section 10.12.5 the results of the model are interpreted in qualitative language.

10.12.2 Grammatical effects in AVCS

Clausal effects have for the most part been ignored, dismissed or controlled for by researchers of AVCS. For instance, in his study De Sutter (2009) only includes subordinate clauses and only when they begin with the complementizer dat. As such, clausal effects cannot be properly evaluated, as there has been no variability to examine. Be that as it may, they have not gone unnoticed for the Latin PAC alternation, even if the research thus far has only been qualitative in nature. Clausal environments in which it has been claimed the two serializations are favoured/disfavoured are reported below.

10.12.2.1 Relative clauses

Devine & Stephens (2006: 184) report that there is “widespread raising of the auxiliary” in relative clauses. Möbitz (1923: 125) also notes that relative clauses provide a favouring environment for est factus serializations (cf. (32)).

(32) ex pellibus quibus erant tectae naues
from skin.F.PL.ABL REL.F.PL.ABL be.3PL.IMPF.IND cover.PTCP.LPRF.F.PL.NOM ship.F.PL.NOM

‘from the hides with which the ships were/had been covered’ (Caes. Civ. 3. 15. 3, cf. Devine & Stephens 2006: 184, ex. 89)

10.12.2.2 Concessive clauses

Vogel (1937: 38, 45, 50, 61) observes that concessive clauses provide a favouring habitat for est factus orders, and the fronting of esse more generally (cf. (33)).

(33) etsi totius diei continenti labore erant confecti
although whole.M.SG.GEN day.M.SG.GEN continuous.M.SG.ABL labour.M.SG.ABL be.3PL.IMPF.IND wear.out.PTCP.LPRF.M.PL.NOM
‘although they were/had been worn out as a result of uninterrupted toil which lasted the entire day’ (Caes. Civ. 3. 97. 4)

10.12.2.3 Temporal clauses

Temporal clauses show variable behaviour depending on the nature of the complementizer. Devine & Stephens (2006: 186–187) note that “[f]or cum, while there are some occurrences with postverbal auxiliary, the preverbal auxiliary predominates” (cf. (34a)), whereas “with ubi, while there are some instances of preverbal auxiliary, the postverbal auxiliary predominates” (cf. (34b)).

(34) a. diu cum esset pugnatum
for.a.long.time when be.3SG.IMPF.SBJV fight.PTCP.LPRF.N.SG.NOM
‘when fighting had taken place for a while’ (Caes. Gal. 1. 26, Devine & Stephens 2006: 186)

b. ubi eo uentum est
when thither come.PTCP.LPRF.N.SG.NOM be.3SG.PRS.IND
‘when they had come to that place’ (Caes. Gal. 1. 43 Devine & Stephens 2006: 187)

For ubi-specific effects, see also Vogel (1937: 49, 55, 69).
It should be noted that there are many other clause types than just the three mentioned here, not least main clause declaratives and interrogatives. In the next sections, evidence will be given for how these contribute to PAC responses.

10.12.3 Variable profile and operationalization

A predictor variable is configured for the clause type in which the PAC is situated (ClauseType). Specifically, this is a factor with \( k = 14 \) levels.\(^{35}\) The following clause types are recognized in this annotation scheme: main clause declaratives (declarative), main and subordinated interrogatives (interrogative), accusative and infinitive clauses (aci), accusative and infinitive clauses involving a relative (aci-relative), nominative and infinitive clauses (nci), cum clauses (cum), ubi temporal clauses (ubi-temporal), other temporal clauses (other-temporal), conditional clauses (conditional), causal clauses (causal), relative clauses (relative), comparative clauses (comparative), result and purpose ut/ne/quin-clauses (ut-clause), and concessive clauses (concessive).

10.12.4 Statistical modelling

<table>
<thead>
<tr>
<th></th>
<th>est factus (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>declarative</td>
<td>300 (32.54)</td>
<td>622 (67.46)</td>
<td>922</td>
</tr>
<tr>
<td>ACI</td>
<td>122 (50.62)</td>
<td>119 (49.38)</td>
<td>241</td>
</tr>
<tr>
<td>cum</td>
<td>95 (48.97)</td>
<td>99 (51.03)</td>
<td>194</td>
</tr>
<tr>
<td>other-temporal</td>
<td>27 (42.19)</td>
<td>37 (57.81)</td>
<td>64</td>
</tr>
<tr>
<td>interrogative</td>
<td>33 (33.33)</td>
<td>66 (66.67)</td>
<td>99</td>
</tr>
<tr>
<td>NCI</td>
<td>4 (21.05)</td>
<td>15 (78.95)</td>
<td>19</td>
</tr>
<tr>
<td>conditional</td>
<td>38 (40.86)</td>
<td>55 (59.14)</td>
<td>93</td>
</tr>
<tr>
<td>causal</td>
<td>45 (48.91)</td>
<td>47 (51.09)</td>
<td>92</td>
</tr>
<tr>
<td>relative</td>
<td>219 (43.45)</td>
<td>285 (56.55)</td>
<td>504</td>
</tr>
<tr>
<td>comparative</td>
<td>29 (39.73)</td>
<td>44 (60.27)</td>
<td>73</td>
</tr>
<tr>
<td>ubi-temporal</td>
<td>1 (4.76)</td>
<td>20 (95.24)</td>
<td>21</td>
</tr>
<tr>
<td>ut-clause</td>
<td>36 (52.94)</td>
<td>32 (47.06)</td>
<td>68</td>
</tr>
<tr>
<td>concessive</td>
<td>9 (90.00)</td>
<td>1 (10.00)</td>
<td>10</td>
</tr>
<tr>
<td>ACI-relative</td>
<td>1 (11.11)</td>
<td>8 (88.89)</td>
<td>9</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
<td>2409</td>
</tr>
</tbody>
</table>

Table 10.44: Relationship of clause type and PAC serialization

For this predictor, a single multilevel model was set up, because of the great number of factor levels resulting in model convergence problems when random slopes are included. Specifically, this model includes only random intercepts for the verb’s lemma and the text from which the observations were taken.

```r
clause.ri <- glmer(RealizationOfPAC ~ ClauseType +
                   (1 | LemmaParticiple) + (1 | Text), pac, family = binomial)
```

An ANOVA comparing this model with the baseline vcm is reported in Table 10.45. As this ANOVA clearly shows, the addition of the clause type predictor significantly reduces the deviance from 3139.41 to 3084.23, and the model clause.ri has the lower AIC, indicating the inclusion of the predictor does not affect the parsimonious quality of the model.

The parameters for this model are therefore reported in Table 10.46. As usual, the logit probability estimate is in terms of a factus est outcome. The predictor variable is coded with main clause declaratives as the baseline level (= 0), against which the remaining levels are contrasted via treatment coding.

\(^{35}\)I decided not to collapse levels here, because some of the levels with low cell counts have been mentioned in previous literature, and it will be important to corroborate their effects.
The intercept indicates that main clause declaratives have an logit probability of 0.5351 (in odds 1.71), meaning that they occur 63.07% of the time with factus est (it is different from the percentage given in the overall crosstabulation in Table 10.44 because the latter does not take into account lemmatic/textual fluctuations). There are several irrelevant predictor levels: specifically, temporal clauses with a complementizer other than ubi/cum, interrogatives, ncis, conditional clauses, comparative clauses, and ACI clauses involving a relative pronoun do not substantially differ from main clause declaratives as to whether they favour/disfavour factus est. The 95% confidence intervals for all these factor levels all span across zero.

Seven estimates are significantly different from the baseline level of main clause declaratives, with one coefficient being positive (for the level ubi-temporal) indicating a stronger preference for factus est and the rest being negative indicating a stronger preference for est factus. Let us now run through these coefficients in terms of decreasing logits. First, when the clause is an ubi temporal clause as opposed to when it is a main clause declarative, the logit probability of a factus est outcome increases by 2.4224 (an odds ratio of 11.27), and the confidence intervals associated with this estimate lie between 0.3232203 and 4.52148058 on the logit scale. This is in the direction claimed by the previous literature. Second, when the clause is a relative clause as opposed to when it is a main clause declarative, the logit probability of a factus est outcome decreases by $-0.4851$ (an odds ratio of 0.61564141), and the confidence intervals associated with this estimate lie between -0.7268366 and -0.24334465 on the logit scale. This is in the direction claimed by the previous literature. Third, when the clause is a cum-clause as opposed to when it is a main clause declarative, the logit probability of a factus est outcome decreases by $-0.6502$ (an odds ratio of 0.52), and the confidence intervals associated with this estimate lie between -0.9849157 and -0.31540401 on the logit scale. This is in the direction claimed by the previous literature. Fourth, when the clause is a causal clause as opposed to when it is a main clause declarative, the logit probability of a factus est outcome decreases by $-0.7070$ (an odds ratio of 0.49), and the confidence intervals associated with this estimate lie between -1.1609254 and -0.25313106 on the logit scale. Fifth, when the clause is an ut-clause as opposed to when it is a main clause declarative, the logit probability of a factus est outcome decreases by $-0.7847$ (an odds ratio of 0.46), and the confidence intervals associated with this estimate lie between -1.3091987 and -0.26014221 on the logit scale. Sixth, when the clause is an ACI clause as opposed to when it is a main clause declarative, the logit probability of a factus est outcome decreases by $-0.9417$ (an odds ratio of 0.39), and the confidence intervals associated with this estimate lie between -1.2555357 -0.62784993 on the logit scale. Finally, when the clause is a concessive clause as opposed to when it is a main clause declarative, the logit probability of a factus est outcome decreases by $-3.3527$ (an odds ratio of 0.03), and the confidence intervals associated with this estimate lie between -5.5485851 -1.15685486 on the logit scale. This is in the direction claimed by the previous literature.

10.12.5 Discussion

It is of note that several of the factor levels behave as suggested in the previous literature: for instance, ubi-clauses prefer factus est, whilst relative clause, cum-clauses, and concessive clauses strongly repel it. Several new effects have been identified, with causal, ut clauses, and ACI all disfavouring factus est sequences.

Additionally, a hierarchy of probabilistic tendencies for the data sample itself can be identified by utilizing the logit probability estimates:
10.13. Conclusion

In this chapter we have motivated the inclusion of grammatical predictors within the information space, both on theoretical grounds and on empirical grounds. Table 10.47 summarizes the effects. Why do the grammatical predictors behave as they do? One striking possibility is that complexity is involved. That is, the participle which is morphologically encoded for a more “default” level — such as singular, masculine, vowel stems, verbal perfect participles, and unprefixed participles — is accessed faster and can be produced more rapidly, thus occurring earlier in the \( \text{PAC} \). However, it is not clear how it would work for deponents: intuitively, one would imagine that deponents constitute the “marked” or more complex category, given that they are non-canonical and less frequent. There is much more to be investigated here, and what is said must remain speculative.

---

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | 0.5351 | 0.1712 | 3.126 | 0.001775 |
| ClauseTypeaci | -0.9417 | 0.1601 | -5.881 | 4.08e-09 |
| ClauseTypecum | -0.6502 | 0.1708 | -3.807 | 0.000141 |
| ClauseTypeother-temporal | -0.3204 | 0.2756 | -1.162 | 0.245054 |
| ClauseTypeteinterrogative | -0.2835 | 0.2411 | -1.176 | 0.239729 |
| ClauseTypenci | 0.3532 | 0.5973 | 0.591 | 0.554357 |
| ClauseTypetypeconditional | -0.3826 | 0.2367 | -1.617 | 0.002266 |
| ClauseTypetypecausal | -0.7070 | 0.2316 | -3.053 | 0.002266 |
| ClauseTypetyperelative | -0.4851 | 0.1233 | -3.933 | 8.39e-05 |
| ClauseTypetypecomparative | -0.2155 | 0.2778 | -0.776 | 0.437555 |
| ClauseTypetypeubi-temporal | 2.4224 | 1.0710 | 2.262 | 0.023712 |
| ClauseTypetypeut-clause | -0.7847 | 0.2676 | -2.932 | 0.003368 |
| ClauseTypetypeconcessive | -3.3527 | 1.1204 | -2.993 | 0.002767 |
| ClauseTypetypeaci-relative | 0.6668 | 1.0804 | 0.617 | 0.537103 |

Table 10.46: Final glmer for serialization ~ clause type

More nuanced aspects of clause type may be worth considering in later studies. For instance, Devine & Stephens (2006: 190) observe that there is a difference between the \( \text{ut} \)-clauses in \( \text{ut demonstratum est} \) ‘as was pointed out’ and \( \text{erat imperatum} \) ‘as was ordered’, which were treated here as both being comparatives: “the latter compares the event described by the clause with a previous command while the former says that the discourse contains an earlier speech act making the same assertion. The latter is concerned with the equivalence of two properties, the former with the anterior occurrence of a discourse event.”

Finally, it may be worth taking into consideration different types of relative clause (e.g. regular vs. free vs. incorporated relative). For example, it has been pointed out by Vogel (1937: 21) that when the relative constitutes the free object of a higher verb, then a \( \text{factus est} \) response might be stimulated (e.g. \( \text{} \)).

(35) \( \text{ubi} > \text{aci relative} > \text{nci} > \text{main clause declarative} > \text{comparative} > \text{interrogative} > \text{other temporal} > \text{conditional} > \text{relative} > \text{cum} > \text{causal} > \text{ut} > \text{aci} > \text{concessive} \)

The magistrates conceal what seems appropriate’ (Caes. \( \text{Gal.} \) 6. 20)
### Variable Summary of Effect

<table>
<thead>
<tr>
<th>Variable</th>
<th>Summary of Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auxiliary morphology</td>
<td>present indicative &gt; (perfect) &gt; imperfect subjunctive</td>
</tr>
<tr>
<td></td>
<td>&gt; present subjunctive &gt; present infinitive &gt; future</td>
</tr>
<tr>
<td></td>
<td>&gt; pluperfect &gt; imperfect indicative</td>
</tr>
<tr>
<td>Participle category</td>
<td>perfect &gt; (perfect adjectival) &gt; (future) &gt; perfect ambiguous &gt; gerundive</td>
</tr>
<tr>
<td>Participle prefixation</td>
<td>unprefixed &gt; prefixed</td>
</tr>
<tr>
<td>Impersonality</td>
<td>—</td>
</tr>
<tr>
<td>Deponency</td>
<td>deponent &gt; non-deponent</td>
</tr>
<tr>
<td>Conjugational Class</td>
<td>vowel stem &gt; consonant stem</td>
</tr>
<tr>
<td>Person</td>
<td>—</td>
</tr>
<tr>
<td>Number</td>
<td>singular &gt; plural</td>
</tr>
<tr>
<td>Gender</td>
<td>masculine &gt; (neuter) &gt; feminine</td>
</tr>
<tr>
<td>Clause type</td>
<td>ubi &gt; (ACI relative) &gt; (NCI)</td>
</tr>
<tr>
<td></td>
<td>&gt; main clause declarative &gt; (comparative)</td>
</tr>
<tr>
<td></td>
<td>&gt; (interrogative) &gt; (other temporal) &gt; (conditional)</td>
</tr>
<tr>
<td></td>
<td>&gt; relative &gt; cum &gt; causal</td>
</tr>
<tr>
<td></td>
<td>&gt; ut &gt; ACI &gt; concessive</td>
</tr>
</tbody>
</table>

Table 10.47: Summary of effects of grammatical predictors (at $p < 0.05$), showing alignment which different orders. Bold indicates the reference level; parentheses indicate non-significance compared to the baseline.
Chapter 11

Semantic and Pragmatic Predictors

11.1 Introduction

In this chapter I discuss the interface between syntactic variability and semantics and pragmatics. The verbal semantics of the predicate will be examined in Section 11.2, followed by a discussion of the potential effect of animacy in Section 11.3. In Section 11.4 a set of pragmatic predictors are motivated concerning information structure. Overall, we find strong and reliable effects in this domain, with the exception of animacy.

11.2 Verbal semantics

In this section, I discuss a central concern of the pac alternation, namely its verbal semantics, which have not yet been analyzed in detail in the literature. The focus is on the notions of eventuality aspect (also Aktionsart, e.g. Vendler 1957) and verb class (e.g. Levin 1993).

11.2.1 Eventuality aspect

Before we begin, there are several related notions that fall under the head of aspect, that we need to distinguish. First, aspect is distinct from a notion we have already dealt with, namely tense: while both tense and aspect are concerned with time in a general sense, specifically tense on the one hand locates an eventuality to a reference point and is thus deictic, whereas aspect on the other deals with an eventuality’s “internal temporal constitution” (Comrie 1976: 52, cf. 3). Comrie (1976: 3) terms the former “situation-external time”, the latter “situation-internal time”. It is the latter that concerns us.

Second, aspect partitions into two different types, grammatical aspect and eventuality aspect. The former type of aspect looks at whether a situation is perceived as a complete whole (“perfective”) or as ongoing (“imperfective”): for instance, John walked the dog indicates that the eventuality of John walking the dog is viewed as complete, whereas John was walking the dog focuses on the eventuality as going on through time and, crucially, as unfinished.¹ This information is typically expressed grammatically, and is thus known as “grammatical aspect”.² Eventuality aspect, as I will call it, looks at the inherent aspectual contours of lexical verbs and the arguments and adjuncts that they combine with. For instance, while John is running can potentially go on indefinitely, John is running the shop can only go on until John has reached the shop, and at that stage it terminates. This type of aspect is typically concerned with whether there is a temporal bound specified in the verb’s lexeme or in the context via arguments or adjuncts, whether it inherently involves duration or no duration, and whether it involves change or no

¹Note that the eventuality in both is positioned deictically in the past relative to the time of utterance. Thus, the semantic tense is the same.
²This type of aspect is also known as “viewpoint aspect” (Smith, 1991).
change. Other terms for it are “lexical aspect” and “Aktionsart” (see Sasse 2002 for other terms and a review).

In terms of representation, in a layered framework, e.g., that of de Swart (2012), where the semantics is mapped onto a syntax, grammatical aspect targets the IP ("Inflectional Phrase"), and eventuality aspect targets the VP ("Verb Phrase"), the core verbal projection. It is assumed that tense — the deictic temporal information — is hierarchically higher than both. See (1).

![Diagram]

3

11.2.1.1 Classes and features

Here, we explore the linguistic details of eventuality aspect from the perspective of English, examining the linguistic features that determine this type of aspect and the classes of eventuality that result from these properties.

Following the seminal work of Vendler (1957), it is usual to classify eventualities into four types: states are distinguished from events, and there are three types of event, namely activities, accomplishments, and achievements. Some typical English examples of each type are given in (2)–(5).

(2) States: knowing Latin, believing in god, existing, involving complications, living in Balham, liking custard

(3) Activities: running, vibrating, playing, fighting, dancing, speaking

(4) Accomplishments: eating a slice of cake, drawing a picture, running a mile, shaving, cooking, reading a speech

(5) Achievements: arriving at school, coughing, dying, finding a brooch, leaving a restaurant, reaching the top of a mountain

Researchers interested in eventuality types such as the above have sought to find what featural primitives distinguish each type. Three features are at the fore here — dynamism, telicity, and durativity (see e.g. Smith 1991). 6

1. The presence or absence of dynamism is the defining feature in the distinction between states on the one hand and events on the other. There are several sub-features pertinent to the distinction — homogeneity and energization.

(a) A predicate is non-dynamic if it is both upwardly homogeneous and downwardly homogeneous to moments. 7 A predicate is upwardly homogeneous if, when \( x = P \) and \( y = P \), it is true that the sum of \( x \) and \( y \) also equals \( P \). To exemplify, in John knows French, the combination of John knows French and John knows French sum to John knows French, assuming constancy of reference. The sentence John ran is also upwardly homogeneous: two intervals in John ran and John ran sum to John ran, again assuming constancy of reference.

---

5Given that in some theories the external argument is projected within the vP, one might want to include this here as well.

4The distinction is based on Aristotelian notions of energesia and kinesis.

6Other types have been suggested; see below.

By contrast, *John built a house* is not upwardly homogeneous: the sum of *John built a house* and *John built a house* is not *John built a house*, but *John built two houses*.\(^8\) A predicate is downwardly homogeneous if, for moments \(x\) and \(y\) in \(P\), where \(x \in P\) and \(y \in P\), it is the true that \(x = P\) and \(y = P\) and \(x \prec y = P\). To exemplify, *John knows French* is true however small we divide it up. To assert that John knew French from 2000 to 2013 is also the same thing as asserting that John knew French at 11:58:45 on February 3, 2008. By contrast, *John ran from 2pm to 4pm yesterday afternoon* is not so analyzable. If I assert that John ran from 2pm to 4pm yesterday afternoon, it is not literally the case that he was running at 3:13:45. He might at that very moment have been kicking his left foot forward and moving his right arm back or he might have stopped off at Waitrose to pick up a bottle of water as a refreshment and be handing cash to the shop assistant — this does not literally constitute running. Thus, whereas states are downwardly homogeneous down to moments, events, even if they are upwardly homogeneous, are not downwardly homogeneous to moments.

(b) Second is energization. Comrie (1976: 49) sums this up: “With a state, unless something happens to change that state, then the state will continue... With a dynamic situation, on the other hand, the situation will only continue if it is continually subject to a new input of energy... To remain in a state requires no effort, whereas to remain in a dynamic situation does require effort”. In the case of *John is walking in the park*, for John to remain in the eventuality of walking in the park he has to continually put energy into moving his legs in a way that constitutes walking. But to remain in the state of knowing French, John has to do nothing at all. He may cease to know French (i.e. forget it), but that is a change of state (an event), and not a state per se.

Events (activities, accomplishments and achievements) are [+dynamic], since they are not both upwardly and downwardly homogeneous and are energized, whereas states are [−dynamic], since they are both upwardly and downwardly homogeneous and not energized.

2. Telicity is whether an eventuality is conceived of as having a “natural” or “inherent endpoint” or not. For example, *John baked a cake* has a natural endpoint: the eventuality of baking the cake is over when the cake is baked and cannot continue past that point. If he were to bake the cake again, then that would be another event of baking. Similarly, in *John ran to the shop* the eventuality inherently terminates when John is at the shop. These are predicates that have a “clearly defined resultant state”. In contrast, *John ran* has no natural endpoint: it can continue for as long as John can sustain his running. Similarly, *John loves Mary* also has no natural endpoint: it can go on indefinitely. Eventualities that have an inherent endpoint are telic [+telic], whereas those that do not are atelic, in feature terms [−telic]. A predicate is atelic if it is upwardly homogeneous.

3. Durativity is “one of the key aspectual properties of situations” (Smith, 1991), and concerns the extent to which the eventuality being described is conceived of as extending through time or as happening instantaneously. For example, *John built the house* is conceived of as extending through time, because it takes John time to reach the culmination of building the house; there are several stages involved building a house before it is built. By contrast, achievement predicates such as *John reached the top of the mountain* or *John coughed* involve no time at all: they happen in an instant. Eventualities that extend through time are durative, [+durative], whereas those that happen instantaneously are punctual, [−durative]. It should be noted that the notion of “instantaneity” is an idealization, as has been remarked by several aspectologists. Comrie (1976) observes that “the single act of coughing... is not punctual in the strict sense, but rather refers to a situation that lasts for a very short time”. Similarly, Smith (1991) notes “an instantaneous event may take milliseconds, perhaps even enough time to be perceptible, without marring its categorization as

---

\(^8\)Note, however, that *John built houses + John built houses = John built houses*. 

199
CHAPTER 11. SEMANTIC AND PRAGMATIC PREDICTORS

<table>
<thead>
<tr>
<th></th>
<th>Dynamic</th>
<th>Telic</th>
<th>Durative</th>
</tr>
</thead>
<tbody>
<tr>
<td>States</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Activities</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Accomplishments</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Achievements</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
</tbody>
</table>

Table 11.1: Eventuality types and features

“−Durative”]. Durativity separates states, activities and accomplishments from achievements: the former are [+durative], whereas the latter are [−durative].

The above-mentioned features characterize the eventuality aspect types in the following way. States are regarded as being [−dynamic], [−telic], and [+durative]. Activities are [+dynamic], [−telic], and [+durative]. Accomplishments are [+dynamic], [+telic], and [+durative]. Achievements are [+dynamic], [+telic], and [−durative]. This featural information is summarized in table 11.1.9

11.2.1.2 Lexical and constructional fluctuations

11.2.1.2.1 Telicizing telics Inherently atelic eventualities — states and activities — can be telicized via coercion in certain discourse settings. For states, this involves reinterpretation as an inchoative, where the state is asserted to come into being, as exemplified in (6). Here the state has been coerced into an achievement.

(6) The sun went down, and in no time at all the sky was black
  (= the sky went black)

For activities, the coercion may involve reinterpretation as an achievement with inchoative semantics, as in (7). Alternatively, it may receive a ‘job-to-be-done’ interpretation, in which case the eventuality is coerced into an accomplishment. For instance, in (8a), it may be that it is John’s job to sing a pre-determined song: the eventuality comes to an end when the song is sung. Similarly, in (8b), it might have been John’s job to run a 100m race: the eventuality comes to an end when he reaches the finish line. This reading is particularly clear when the quantized object is pragmatically available in the discourse context.

(7) John ran at 2pm
  (= John began a running event at 2pm)

(8) a. It took John two minutes to sing
    (= John sang a pre-specified song)

b. John ran in 14 seconds
    (= John ran a 100m race)

11.2.1.2.2 Multiple events In certain contexts telic eventualities, such as sending ambassadors to Rome, can be used in “multiple event” contexts. Smith (1991: 85) defines these as “[consisting] of an unbroken series of sub-events which are conceptualized as a single, ongoing event”. Put differently, they consist of a series of microevents \{x, y, z\} in each of which a single event takes place; when they are summed together \(x + y + z\) they make up a macroevent \(X\). In this way, telic eventualities are thus rendered atelic, as they are upwardly homogeneous, and become activities. Example (9) refers to a single telic event, in

9In addition, Smith (1991), amongst others, have proposed that a fifth category should be established for “semelfactives” such as cough and knock. Smith (1991) argues that they have a feature matrix corresponding to our [+dynamic], [−telic], and [−durative]. Other semanticists include them within the achievement class. In this thesis, we shall not follow Smith (1991). My reasoning for this is that while Smith (1991) classifies them as [−telic] it is not clear whether this is the case. For instance, a single event of coughing (one cough) necessarily comes to an end once the cough is over. In this respect, some semanticists call them “full-cycle resettable” (Filip 2012 quoting Talmy 1985).
11.2. VERBAL SEMANTICS

which Crassus sent ambassadors to Rome en masse. By contrast, with the insertion of a for-adverb of duration, the eventuality is coerced into an activity, as in (10), which asserts that Crassus spent three hours iteratively sending ambassadors to Rome.

(9) Crassus sent ambassadors to Rome at noon
(10) Crassus sent ambassadors to Rome for three hours

11.2.1.2.3 Stative and eventive ambiguities There are also fluctuations with respect to whether predicates can be interpreted statively or eventively. Constructionally, it has been argued that perfects, generics, habituals, and adjectival passives can stativize the predicate within a construction (Rothmayr, 2009: 35–36), but there are also ambiguities at the lexical level too. Specifically, drawing on earlier literature, Rothmayr (2009) identifies predicates that show lexical ambiguities: these include verbs that participate in the instrumental alternation, object-experiencer predicates, dispositional predicates, threaten-type predicates, and perception verbs. The examples in (11) demonstrate this for instrumental alternation: in (11a) there is an event in which the cake is incrementally covered in icing until it is completely covered, whereas in (11b) the cake just happens to be that way.

(11) a. John covered the cake with icing
b. Icing covered the cake

This distinction is pertinent to Latin too, where the same predicate can have stative/eventive ambiguities: “[i]n addition to their regular eventive readings, some change of state verbs like cingo ‘surround,’ munio ‘fortify,’ claudio ‘close off,’ adorno ‘embellish,’ also have purely stative readings in which the copresence of an inanimate entity causes some state (without the usual process subevent)” (Devine & Stephens, 2013: 119). Contrast, for example, amnis urbem muniuit ‘the river protected the city’ with Caesar urbem muniuit ‘Caesar protected the city’. In the former the predicate receives a stative interpretation, in the latter an eventive interpretation.

11.2.2 Verb class

In addition to eventuality structure based on properties of dynamism, telicity, and duration, verbs behave differently syntactically based on their finer-grained lexical semantic groupings, or their verb class, which take into account thematic roles and argument structure. Levin (1993) diagnoses semantic 57 classes of verbs based on their differing responses to a number of syntactic alternations. For instance, verbs of sending and carrying behave differently to verbs of motion. Although the verb classes are specific to English, as they are based on syntactic criteria, they nevertheless constitute a valuable testing ground for exploring semantic effects.

11.2.3 Verbal semantics and grammatical variability

Semantics are well known to play an important role in (e.g.) the English dative alternation, where verbs of prevention of possession like cost strongly favour the NP-recipient variant (cf. Green, 1971; Oehrle, 1976; Gropen et al., 1989; Pinker, 1989; see Bresnan et al. 2007 for a summary). However, very little work appears to have been done on the relationship of verbal semantics and AVC serialization.

Nevertheless, Devine & Stephens (2006) do assert that semantic differences have an effect on the choice of PAC variant in Latin. They argue that the factus est variant is positively associated with eventualities with an eventive component, and disprefers associating with eventualities with only a stative component (2006: 180): “for the auxiliary to be postverbal [i.e. factus est] the verb... is almost always eventive . . ., rather than a stative property predicated of the subject”. One example they cite later which may be seen
as support for this assertion is the case of *profectus est*, an eventive verb which “almost always has the auxiliary after the participle in main clauses” (2006: 188).

Devine & Stephens (2006: 180) mention another important semantic factor related to this: achievement predicates, which are telic and “often inherently punctual”, are strongly associated with the *factus est* variant, as in the following example. There seem to be important distinctions within the eventive class itself, then.

(12) ubi *signum datum sit*, clamorem
when signal.N.SG.NOM give.PTCP.L.PRF.N.SG.NOM be.3SG.PRS.SBJV shout.M.SG.ACC
ommnes tollere iubet
all.M.PL.ACC raise.PRS.IND order.3SG.PRS.IND
‘When the signal was given, he bade them all to raise a clamour’ (Liv. 3. 28. 2)

Three points are worth noting. First, it needs to be stressed that Devine & Stephens do not assert that the effect of the verb’s semantics is completely categorical. For instance, they have examples that clearly show that certain types of subordinate clauses and preverbal focus can override the eventive construal of the verb, resulting in an *est factus* variant, e.g. the unde-clause in *unde erant profecti* (Caes. Gal. 1. 28) and focus on the temporal adjunct *paucis post diebus est profectus* (Caes. Civ. 3. 33). Second, in their recent book on semantics in Latin, Devine & Stephens (2013: 150) report a number of examples of *pac* with strong stative readings, so strong in fact that “[a]gentive-eventive readings would give nonsense”. The three examples they report are in the *factus est* variant. It would have been nice if they had chosen examples that would have fitted in with their 2006 hypothesis! Finally, nothing has been suggested about the extent to which different lexical semantic verb classes bear on the alternation. In this study, I shall seek to provide quantitative evidence for the aktionsart claims of Devine & Stephens (2006) and add to our lexical semantic knowledge of the construction by including a predictor for the PAC’s verb class.

### 11.2.4 Variable profile and operationalization

#### 11.2.4.1 Eventuality aspect

A variable is annotated for the eventuality aspect of the predicate (*EventAspect*), and is distinguished for four levels: noneventive (stative, non-dynamic predicates), activity (dynamic, durative, and atelic predicates), accomplishment (dynamic, durative, telic predicates), and achievement (dynamic, punctual, telic predicates).

This variable is not straightforward to annotate, especially as we are dealing with a dead language for which there are no diagnostic tests as regards eventuality aspect. What will be done, therefore, is to apply well-established diagnostics to the meaning of the PAC in its discourse context. For instance, Devine & Stephens (2013), who discuss eventuality aspect in detail for Latin, appear to assume that the diagnostic of delimitation and containment adverbials work just as well for Latin as they do for English. For instance, they say that “a durative (time-how-long) adverbial can explicitly delimit an activity”, and they give examples of the predicate *pugno* which can only be classified as an activity based on its translation ‘fight’. Therefore, I am not alone in this regard. While this is not foolproof, applied sensitively and with important regard to a predicate’s decompositional semantics, it does allow us to illustrate how the variable might behave in a fairly “rough-and-ready” way. In what follows, I will therefore set out the most influential tests for diagnosing a predicate’s category, based on Dowty (1979), the source of choice in this regard.

#### 11.2.4.1.1 Dynamic vs. non-dynamic

The following set of diagnostics set stative eventualities apart from eventive ones.

---

10The crosslinguistic validity of these is uncertain, however.
11.2. VERBAL SEMANTICS

11.2.4.1.1 Pseudo-clefts  As we saw earlier, events happen and thus advance the narrative, whereas states obtain and typically present backgrounded information. Thus, one diagnostic to distinguish the two is the \textit{What happened (next) was...} pseudo-cleft test.

(13) a. *What happened (next) was John loved Mary (state)
    b. What happened (next) was John ran (activity)
    c. What happened (next) was John read the letter (accomplishment)
    d. What happened (next) was John died (achievement)

Similarly, the pseudo-cleft of the type \textit{what NP did} also diagnoses between states and events.

(14) a. *What John did was believe in ghosts
    b. What John did was run
    c. What John did was read the letter
    d. What John did was lose his glasses

11.2.4.1.2 Simple present  An influential diagnostic is the use of the simple present. In clauses in the simple present tense (e.g. \textit{John knows, runs, etc.}), stative eventualities are true at the utterance time, whereas events are not necessarily true at the utterance time. For instance, if I utter the sentence \textit{John knows what aspect is}, then I am asserting that it is true the moment I say it. By contrast, if I utter the sentence \textit{John runs}, then I am not asserting that at the present moment he is running (not necessarily true that he is running at that particular moment). In such cases, events have a habitual interpretation, whereas states refer to the present moment of utterance. Thus, Kearns (2000: 210) observes “Non-state sentences . . . are generally interpreted as habitual in the simple present tense . . . . States are interpreted as holding at the time of utterance, whether they are enduring states or temporary situations”.

(15) a. John knows French
    b. John cycles (habitualized activity = state)
    c. John walks to work (habitualized accomplishment = state)
    d. John arrives at work on time (habitualized achievement = state)

11.2.4.1.3 Counterfactuals Rothstein (2004: 16), following Vendler (1957: 148), points out that stative eventualities are bad in counterfactuals of the type \textit{Even if NP could P, NP wouldn’t}. Rothstein (2004) observes that this would lead to a contradiction for states, where \textit{would} is entailed by \textit{could}.

(16) a. *Even if John could know the answer, he wouldn’t (state)
    b. Even if John could run, he wouldn’t (activity)
    c. Even if John could build a house, he wouldn’t (accomplishment)
    d. Even if John could find his watch, he wouldn’t (achievement)

Note, however, that if the stative eventuality seems acceptable in such counterfactual constructions, this is because it has shifted into an event (see above). For instance, the sentence \textit{Even if John could own a house, he wouldn’t}, is not stative, even though \textit{own} is generally used in stative predications. Here, the predication involving \textit{own} is coerced into an achievement, in meaning roughly equivalent to ‘come to own’, ‘acquire’, predicates which are inherently eventive.

11.2.4.1.2 Telicity  A number of diagnostics have been proposed that separate telic eventualities from atelic ones. The telicity tests make an overt reference to the endpoint of an eventuality. Given that telic eventualities inherently have an endpoint, such predicates pass telicity tests with ease. This is not the case with atelic eventualities: the telicity tests are making reference to an endpoint, when in fact atelics
do not have such an endpoint. Thus, they are incompatible with atelics per se, as they give rise to a contradiction — i.e. specification of an endpoint when there isn’t one.

11.2.4.1.2.1 Container adverbials  Telic eventualities naturally take container adverbials that express ‘time-within-which’, such as in five minutes, in an hour, in a week. With accomplishments they modify the duration of the eventuality before its resultant state comes into being, while with achievements they express how much time it takes before the instantaneous change of state takes place. Thus, in (17a) the adverbial expresses that building was taking place for a week before something that represented a house came into an effected state, while in (17b), it expresses that one week elapsed from some unspecified reference point (in this example, maybe when he lost the watch, maybe when he’d bought another one as a replacement) until he found it again. For Devine & Stephens (2013), the former is the “inner reading” while the latter is the “outer reading”.

(17) a. John built a house in a week (accomplishment)
    b. John found his watch in a week (achievement)

By contrast, atelic eventualities — states and activities — are not possible with in-adverbials, because they have no intrinsic culmination point that can be reached.

(18) a. *John knew the answer in a day (state)
    b. *John ran in an hour (activity)

If they do seem acceptable, this is because they have been coerced into achievements or accomplishments.11

(19) a. John knew (i.e. came to know) the answer in a day (telicized state = achievement)
    b. John ran (i.e. began the onset of running) in an hour (telicized activity = achievement)
    c. John ran (i.e. ran an understood distance) in an hour (telicized activity = accomplishment)

11.2.4.1.2.2 Take x time  An alternative to the above test is the take x time test, which again is permissible only with telic eventualities and atelic eventualities that are pragmatically telicized.

(20) a. It took John a day to know the answer in a day (state)
    b. It took John an hour to run (activity)
    c. It took John a week to build a house (accomplishment)
    d. It took John a week to find his watch (achievement)
    e. It took John a day to know (i.e. come to know) the answer in a day (telicized state = achievement)
    f. It took John an hour to run (i.e run an understood distance) (telicized activity = accomplishment)

11.2.4.1.2.3 Imperfective paradox  Atelic predicates such as activities are upwardly homogeneous. As a result, one can conclude from a sentence John is walking that John has walked. However, telic predicates such as accomplishments and achievements are not homogeneous. Consequently, one cannot conclude from the accomplishment eventuality John is making the cake that John has made the cake. Similarly, one cannot conclude from the achievement eventuality John is dying that John has died. This is known as the “imperfective paradox” (Dowty, 1979: 133). The test that gives rise to the imperfective paradox is formalized by Dowty (1979: 57) in the following way: “If $\phi$ is an activity verb, then $x$ is (now)
11.2 VERBAL SEMANTICS

φing entails that x had φed. If φ is an accomplishment verb, then x is (now) φing entails that x has not (yet) φed”. Note that this test only works well if referents are kept constant, as pointed out by Dowty (1979).

11.2.4.1.3 Duration

11.2.4.1.3.1 For-adverbials Atelic eventualities which are durative are natural with adverbials of delimitation, headed by e.g. for. They delimit the span of the eventuality. John’s belief in gods lasted ten years. John’s running lasted an hour.

(21) a. John believed in gods for ten years
    b. John ran for an hour

Telic eventualities are not possible with for-adverbials in their canonical reading.

(22) a. *John built a house for a week
    b. *John rang the bell once for an hour

There are two contexts in which for-adverbials may become acceptable, as has been pointed out by e.g. Dowty (1979); Binnick (1991); Devine & Stephens (2013). In such a case, they have two readings. Consider the following example.

(23) The sheriff of Nottingham jailed Robin Hood for four years

As Dowty (1979: 58)’s example shows, the adverbial phrase for four years can modify the resultant state (that the sheriff of Nottingham put Robin Hood in jail to serve a four year sentence) or the activity phase with an iterated reading, in which the sheriff of Nottingham repeatedly put Robin Hood in jail during a delimited time span of four years.

Similarly with John came for half an hour does not indicate the duration of the coming, but its resultant state and seems elliptical: e.g. John came and stayed for half an hour.

11.2.4.1.3.2 Spend x time Punctual events are incompatible in phrases of the type spend x time doing, whereas activities and accomplishments are fine.

(24) a. John spent five hours reading the newspaper
    b. John spent ten weeks building a model aeroplane
    c. *John spent an hour arriving at the station

11.2.4.1.3.3 Complementation after begin/stop Complementation after begin and stop are possible with activities and accomplishment, because they are durative and dynamic. Achievements are not felicitous with begin or stop, because they are inherently punctual.

(25) a. John began/stopped loving Mary
    b. John began/stopped running
    c. John began/stopped reading the letter
    d. *John began/stopped losing his glasses

Note however that activities and accomplishments have different entailments in the complementation pattern with stop. If John stops running, it is still true that he ran. However, if John stops building a house, it is not true that he built a house; in fact, it suggests that there was some interruption in the process subevent that prevented the resultant state from being effected.
11.2.4.1.3.4 Entailments Finally, Dowty (1979: 59) observes the following: “If John painted a picture in an hour is true, then it is true that John was painting a picture during that hour. But from the truth of [John noticed the painting in a few minutes] it does not follow that John was noticing the painting throughout the period of a few minutes”. Dowty formalizes this in the following way: “If \( \phi \) is an accomplishment verb, then \( x \text{ \&} \phi \text{\&} y \text{ time} \) entails \( x \text{ \&} \phi \text{\&}ing \text{\&} y \text{ time} \). If \( \phi \) is an achievement verb, then \( x \text{ \&} \phi \text{\&} y \text{ time} \) does not entail \( x \text{ \&} \phi \text{\&}ing \text{\&} y \text{ time} \)”. 

11.2.4.1.4 Accomplishments A number of tests separate accomplishments from the other two eventive classes.

11.2.4.1.4.1 Complementation after finish Complementation after finish is only felicitous for accomplishments predicates, which inherently involves a lead up (a process) and a point at which the event naturally ends (Kearns, 2000: 214). This is not possible with activities, because they cannot be finished (there is no natural endpoint to be finished off). It’s also not possible with punctual eventives, because they have no inherent process subevent leading up to the culmination.

(26) a. *John finished running 
b. John finished reading the letter 
c. *John finished losing his glasses

11.2.4.1.4.2 Ambiguity with almost Accomplishments, unlike other eventives, are ambiguous with the adverbial modifier almost. With activities and achievements almost gives rise to a reading whereby the eventuality never took place. For instance, from John almost ran one cannot conclude John did some running; he never got round to running in the first place. By contrast, with an accomplishment predicate the conclusion of John did some painting from John almost painted the picture it not necessarily false, though it may be; for instance, he might have run out of paint before the picture could reach the resultant state of having been painted. Alternatively, he never got round to starting the painting in the first place.

11.2.4.1.5 Additional comments In addition to the above, it was also decided to code PAC observations as being noneventive when the auxiliary has full semantic and stative force (or at least can be so interpreted): therefore, all true adjectives, ambiguous adjectival passives, gerundives, and future participles were so coded. Additionally, if the verb was of a type that allowed both a stative and eventive interpretation (see above), it was coded as being noneventive if there was nothing in the overt syntax to indicate its eventiveness.

11.2.4.2 Verb class

The operationalization of verb class (VerbClass) used here is fairly straightforward, that is by running a script over a variable containing the predicate’s meaning and matching it with one of the classes in Levin (1993). It should be pointed out that not all the classes are used. Specifically, only those classes were used when (i) its total number of observations reached \( n = 60 \) (to facilitate statistical comparisons) and when (ii) the meaning of the verb fell only within a single class and was not crosslisted. This is because we want to know how a predicate behaves when it belongs to a specific class with certainty. Coded as such, there are \( k = 11 \) levels for this predictor: verbs of psychological state (psych), verbs of communication (communication), verbs of removing (removing), verbs of change of possession (changeOfPossession), verbs of motion (motion), verbs with predicative complements (predComp), verbs of appearance, disappearance, and occurrence (appearance), verbs of killing (kill), verbs of putting (putting), verbs of sending (sending) and all other verbs that did not fall into any of these classes (other). Table 11.2 gives some examples of each main class.
11.2. VERBAL SEMANTICS

Verb Class Examples
psychological state deterreo, uereor, admiror
communication dico, mando, nuntio
removing aufero, capio, nudo
change of possession potior consequor, do
motion eo, peruenio, proficiscor
predicative complements creo, habeo, nomen
appearance, disappearance, and occurrence morior, nascor, sequor
killing concido, interficio, occido
putting pono, colloco, sero
sending mitto, transporto, trado

Table 11.2: List of verb classes

11.2.5 Statistical modelling

11.2.5.1 Eventuality aspect

Table 11.3 crossclassifies the levels of predictor variable for eventuality aspect with serialization. A $\chi^2$-test of independence indicates that the two variables are not independent of each other ($\chi^2 = 67.023$, $df = 3$, $p = 1.852e-14$). Achievement predicates strongly prefer *factus est* at a rate of 32.65%, whereas non-eventive predicates and constructions disprefer it in proportional terms.

<table>
<thead>
<tr>
<th></th>
<th>factus (%)</th>
<th>factus (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>noneventive</td>
<td>317 (32.65)</td>
<td>654 (67.35)</td>
<td>971</td>
</tr>
<tr>
<td>activity</td>
<td>129 (33.42)</td>
<td>257 (66.58)</td>
<td>386</td>
</tr>
<tr>
<td>accomplishment</td>
<td>83 (42.13)</td>
<td>114 (57.87)</td>
<td>197</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
<td>2409</td>
</tr>
</tbody>
</table>

Table 11.3: Relationship of eventuality aspect and PAC serialization

However, we have not yet taken into account by-lemma variability, nor that of texts. Thus, in line with the majority of predictors in previous chapters, four models are constructed and compared for this predictor. The first model specifies a term for the predictor variable, alongside random intercepts for lemma and text:

```
event.ri<-glmer(RealizationOfPAC~EventAspect+ (1|LemmaParticiple)+(1|Text),pac,family=binomial)
```

The next three models add superstructure. Specifically, in the second model, varying slopes are included for both lemma and text:

```
event.rs.both<-glmer(RealizationOfPAC~EventAspect+ (EventAspect|LemmaParticiple)+(EventAspect|Text),pac,family=binomial)
```

For the third model, varying slopes are included for lemma alone:

```
event.rs.lemma<-glmer(RealizationOfPAC~EventAspect+ (EventAspect|LemmaParticiple)+(1|Text),pac,family=binomial)
```

For the fourth model, varying slopes are included for text only:

```
event.rs.text<-glmer(RealizationOfPAC~EventAspect+ (1|LemmaParticiple)+(EventAspect|Text),pac,family=binomial)
```
It can be discerned from the ANOVA output in Table 11.4, which compares the above four models, that none of the three more complex models add informative information. There are no significant effects for eventuality aspect with respect to lemma ($\chi^2 = 8.9959$, $df = 9$, $p = 0.4377$) nor with respect to text ($\chi^2 = 6.1025$, $df = 9$, $p = 0.7296$). We accept the random intercepts model for further statistical probing.

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>3073.41</td>
<td>3108.13</td>
<td>-1530.70</td>
<td>3061.41</td>
</tr>
<tr>
<td>15</td>
<td>3082.41</td>
<td>3169.22</td>
<td>-1526.21</td>
<td>3052.41</td>
</tr>
<tr>
<td>15</td>
<td>3085.31</td>
<td>3172.11</td>
<td>-1527.65</td>
<td>3055.31</td>
</tr>
<tr>
<td>24</td>
<td>3093.86</td>
<td>3232.75</td>
<td>-1522.93</td>
<td>3045.86</td>
</tr>
</tbody>
</table>

Table 11.4: ANOVA statistics for various glmers containing eventuality aspect information

The ANOVA comparison which compares the random intercepts model against the baseline vcm (in Table 11.5) indicates that the addition of the predictor variable is warranted. There is a significant reduction in deviance of 71.998 between that of 3139.41 for baseline.vcm and that of 3073.41 for event.ri. In sum, eventuality aspect greatly increases the model fit.

| Df  | AIC     | BIC     | logLik  | deviance | Chisq  | Chi Df | Pr(>|Chisq|) |
|-----|---------|---------|---------|----------|--------|--------|----------|
| 3   | 3139.41 | 3156.77 | -1566.70| 3133.41  | -      | -      | 1.0000   |
| 6   | 3073.41 | 3108.13 | -1530.70| 3061.41  | 72.00  | 3      | 0.0000   |

Table 11.5: ANOVA of baseline vcm and glmer containing eventuality aspect information

The parameter estimates of event.ri now deserve to be inspected, which are displayed in Table 11.6. For the response variable, the logit probability estimate is specified in terms of factus est (coded 1), while for the predictor variable the level noneventive is the baseline category (coded 0). It is not the category with the greatest number of observations recorded on it (that is reserved for achievements), but it is of theoretical and empirical interest, because Devine & Stephens (2006) argued for a split between statives and eventives. The intercept says that when a PAC is non-eventive, the logit probability of a factus est outcome is $-0.33$. The other coefficients are to be compared with respect to this value.

All three parameter estimates are significantly different from the baseline level, and all are negative in the direction claimed by the literature. Let us now run through these coefficients in terms of increasing logits. First, when the PAC is formed of an accomplishment predicate as opposed to a non-eventive predicate, the logit probability of a factus est outcome increases by 0.5283 (an odds ratio of 1.696047), and the confidence intervals associated with this estimate lie between 0.1829427 and 0.87363925 on the logit scale. Second, when the PAC is formed of an activity predicate as opposed to a non-eventive predicate, the logit probability of a factus est outcome increases by 0.7534 (an odds ratio of 2.1242400), and the confidence intervals associated with this estimate lie between 0.4650403 and 1.04178794 on the logit scale. Third, when the PAC is formed of an achievement predicate as opposed to a non-eventive predicate, the logit probability of a factus est outcome increases by the largest amount of 0.9460 (equating to an an odds ratio of 2.5753705), and the confidence intervals associated with this estimate lie between 0.7204145 and 1.17157235 on the logit scale.

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|---------|
| (Intercept) | -0.3308 | 0.1741 | -1.900 | 0.05738 |
| EventAspectachievement | 0.9460 | 0.1151 | 8.219 | 2e-16 |
| EventAspectactivity | 0.7534 | 0.1471 | 5.121 | 3.04e-07 |
| EventAspectaccomplishment | 0.5283 | 0.1762 | 2.998 | 0.00272 |

Table 11.6: Final glmer for serialization ∼ eventuality aspect
11.2.5.2 Verb class

We move next to the verb class predictor. The crossclassification of this variable by serialization choice is given in Table 11.7. This table suggests that there is some degree of fluctuation from the mean, depending on the verb class. For instance, while verbs of appearance strongly disprefer being realized in est factus serializations, verbs of putting and sending are fairly evenly distributed between the two word orders. A \( \chi^2 \)-test of independence reveals this distribution to be statistically significant (\( \chi^2 = 33.015, df = 10, p < 0.001 \))

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>appearance</td>
<td>13 (19.40)</td>
<td>54 (80.60)</td>
<td>67</td>
</tr>
<tr>
<td>change of possession</td>
<td>88 (36.67)</td>
<td>152 (63.33)</td>
<td>240</td>
</tr>
<tr>
<td>communication</td>
<td>48 (28.40)</td>
<td>121 (71.60)</td>
<td>169</td>
</tr>
<tr>
<td>killing</td>
<td>27 (44.26)</td>
<td>34 (55.74)</td>
<td>61</td>
</tr>
<tr>
<td>motion</td>
<td>43 (36.44)</td>
<td>75 (63.56)</td>
<td>118</td>
</tr>
<tr>
<td>other</td>
<td>554 (42.19)</td>
<td>759 (57.81)</td>
<td>1313</td>
</tr>
<tr>
<td>predicative complement</td>
<td>42 (35.90)</td>
<td>75 (64.10)</td>
<td>117</td>
</tr>
<tr>
<td>psych</td>
<td>26 (37.14)</td>
<td>44 (62.86)</td>
<td>70</td>
</tr>
<tr>
<td>putting</td>
<td>37 (48.68)</td>
<td>39 (51.32)</td>
<td>76</td>
</tr>
<tr>
<td>removing</td>
<td>32 (41.03)</td>
<td>46 (58.97)</td>
<td>78</td>
</tr>
<tr>
<td>sending</td>
<td>49 (49.00)</td>
<td>51 (51.00)</td>
<td>100</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
<td>2409</td>
</tr>
</tbody>
</table>

Table 11.7: Relationship of verb class and \( \text{pac} \) serialization

As for statistical modelling, it would normally be requisite to build four multilevel models. However, model convergence issues arose when varying slopes were included, even when the number of evaluations was allowed to exceed 10000. As such, one multilevel model is constructed, specifying the predictor term and varying intercepts for lemma and text:

\[
\text{verbclass.ri} \leftarrow \text{glmer(RealizationOfPAC~VerbClass+}(1|\text{LemmaParticiple})+(1|\text{Text}), \text{pac,family=binomial)}
\]

This model is compared by ANOVA with the baseline \( \text{vcm} \), with the output reported in Table 11.8. The model with the predictor included fares significantly better than the baseline model with a reduction in deviance of 32.87, which is highly significant on a \( \chi^2 \)-distribution with \( 13 - 3 \) degrees of freedom. Therefore, verb class appears to have some effect.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.vcm</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>verbclass.ri</td>
<td>13</td>
<td>3126.54</td>
<td>3201.77</td>
<td>-1550.27</td>
<td>3100.54</td>
<td>32.87</td>
<td>10</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

Table 11.8: ANOVA of baseline \( \text{vcm} \) and \( \text{glmer} \) containing verb class information

The nature of the effect can be discerned by inspecting the model’s parameters (Table 11.9). The response variable \( \text{RealizationOfPAC} \) is coded such that the logit probability is defined in terms of a factus est outcome. The predictor variable \( \text{VerbClass} \) is (0-1)-coded with the level \( \text{other} \) as the baseline, against which all the remaining 10 levels are to be compared. As one can see from Table 11.9, two of the levels are influential, while the remaining eight are not. Specifically, when the \( \text{pac} \) is a verb of appearance (e.g. \( \text{nascor} \) ‘be born’, \( \text{morior} \) ‘die’) as opposed to when it is from a verb classified here as \( \text{other} \) (i.e. the majority of verbs), the logit probability of a factus est outcome increases by 1.34; this is equivalent to an odds ratio of 3.836348. The estimate divided by its standard error of 0.16 yields a \( z \)-value greater than 1.96, indicating the highly statistically significant effect of the parameter. The 95% confidence intervals range from 0.67 to 2.02 in logits (or from 1.96 to 7.51 in odds); this says that in 95% of samples, the
estimate would lie within this range. It is also another way of saying that we can be 95% confident that the population estimate $\hat{\beta}$ would lie within this range. Second, when the $\text{PAC}$ is a verb of communication (e.g. *enuntio* ‘announce’), the logit probability of a *factus est* outcome increases by 0.70, equivalent to an odds ratio of 2.01. The 95% confidence intervals lie well above 0 in logits, that is from 0.27 to 1.13 (or from 1.31 to 3.10 in odds), attesting to the robustness of this parameter estimate.

| Estimate      | Std. Error | z value | Pr(>|z|) |
|---------------|------------|---------|----------|
| (Intercept)   | 0.076419   | 0.163593| 0.467    | 0.64041  |
| VerbClassappearance | 1.344521   | 0.342608| 3.924    | 8.7e-05  |
| VerbClasschangeOfPossession | 0.279819   | 0.175023| 1.599    | 0.10988  |
| VerbClasscommunication | 0.700517   | 0.219413| 3.193    | 0.00141  |
| VerbClasskilling   | 0.005586   | 0.395886| 0.014    | 0.98874  |
| VerbClassmotion    | 0.392065   | 0.283271| 1.384    | 0.16634  |
| VerbClasspredComp  | 0.102810   | 0.225861| 0.455    | 0.64897  |
| VerbClasspsych     | 0.057262   | 0.276247| 0.207    | 0.83579  |
| VerbClassputting   | -0.103785  | 0.275953| -0.376   | 0.70684  |
| VerbClassremoving  | 0.214395   | 0.276929| 0.774    | 0.43882  |
| VerbClasssending   | -0.382571  | 0.263282| -1.453   | 0.14620  |

Table 11.9: Final $\text{GLMER}$ for serialization $\sim$ verb class

11.2.6 Discussion

To draw the results of the two verbal semantic predictors together, on the one hand, eventuality aspect declines in the way predicted by Devine & Stephens (2006), with achievement predicates associating most strongly with *factus est* and noneventive/stative predicates and constructions associating least strongly with this order. Second, the new predictor of verb class, previously unmentioned with respect to the $\text{PAC}$ alternation, appears to be important in two respects — verbs of appearance, disappearance, and occurrence and verbs of communication stimulate the use of *factus est*. Even when the verb’s lemma is controlled for (that is, by including it as a varying intercept), there are still significant semantic differences according to the verb’s class.

Sometimes it is fairly obvious why the serial syntax is affected by semantic factors. Devine & Stephens (2006: 81), on Latin ditransitives, point out that how the eventuality is conceptualized might be responsible, in part at least. For *praedam militi dedit* ‘booty to-the-soldiery gave’ “[t]he booty can be seen as going to the soldiers, giving the conceptual ordering theme — recipient”; there is a movement of the theme to the goal. However, the reasons as to why punctual achievement predicates, verbs of appearance/disappearance/occurrence, and verbs of communication pattern with *factus est* are less obvious, as principles of conceptualization and iconity are hardly in play. But they might be.

Some theories of the syntax-semantics interface argue that eventuality notions are phrase structurally encoded (e.g. Ramchand 2008). In terms of decompositional semantics, there is an initiation subevent, a process subevent, and a result state subevent. The initiation phrase corresponds to the onset of an event (the punctual point of departure for the event), the process phrase corresponds to the activity, and the result state phrase corresponds to the final state (cf. (27)).
11.2. VERBAL SEMANTICS

As an example, take *Mary baked the cake*: in *Mary began to bake*, we focus on the initiation subevent; in *Mary is baking*, the eventuality spans the process phase; in *the cake is baked*, the eventuality spans only the result state phase; in *Mary baked the cake*, all of the subevents are spanned. Mary began the process of baking a cake that ended up in the cake being baked. The template makes more or less the right predictions concerning the direction of the data in Table 11.6 with an *achievements* > *activity* > *accomplishment* > *state* hierarchy favouring *factus est*, assuming the auxiliary is “fixed” in some position in the tree (e.g. Proc\(^0\)). Adjectives and other stative expressions are states and it is reasonable to suppose that participles expressing such a feature would be lexicalized lowest in the tree, specifically somewhere in ResP. Activities are next, lexicalized in ProcP, as they denote the process subevent. Achievements are “over as soon as they have begun” (Rothstein, 2004), so even though they have a result state component, it is as if the initiation is the most salient part of an achievement, and may therefore be lexicalized in InitP. One reason that this might be so is that they are easily reconcilable with the empirical facts. They are difficult to combine with *start*, whereas activities, accomplishments, and states are, yielding an inchoative meaning to the sentence. If, e.g., Spec,Init is already filled by the achievement predicate it could not also simulataneously be filled by *start* which necessarily denotes initiation. Accomplishments are less easy to reconcile with this scheme. One could think of them as being lexicalized in between ProcP and ResP, looking both to the process subevent because of their inherent duration and to the result state subevent because of their telicity. However, as we have seen, the nature of the relationship is certainly not categorical, so such a syntax would be required to have a stochastic component. The exact nature of this mapping of the decompositional semantics onto the syntax will need to be examined much further in later research, and I cannot give it the attention it deserves here.

An alternative way to think about the order difference is again in terms of a flat syntax informed by cognitive complexity: canonical eventualities are typically punctual or durative, whereas less typical eventualities are stative. Because they are more canonically eventive, achievements and activities are put earlier, as they are easier to access, whereas statives are less easy to access and put last. However, the accessibility of the different eventuality aspects will need to be examined in psycholinguistic and neurolinguistic research, before such correlations can cleanly be posited, but it is a direction that is worth examining. Alternatively, we may think of it conceptually in a different way: an “urgency-first” principle. Punctual eventualities advance the narrative, statives typically provide background information by predicating a property of the subject, that is, it is more important to put the participle denoting an important event earlier on.

To sum up and reflect, verbal semantics do appear to play a bivariate role on the PAC alternation.
However, there are several key deficits to this analysis that need to be stressed. As noted above, habitualls, generals, and perfect constructions can stativize a sentence. These have not been examined here, because of time. Other ways of operationalizing verbal semantics may be of use which have not been explored here. For instance, the predicate *inuenio* ‘find’ which is strongly punctual (achievement) is found less often in the imperfect form *inueniebat* (7 occurrences) than the perfect counterpart *inuenit* (402 occurrences); by contrast, an activity verb such as *pugno* is realized in the imperfect for *pugnabat* 22 times compared with 46 for the perfect form *pugnauit*. Taking the ratio of imperfect to perfect forms may be a quick and approximate way of assessing the value of this aspect on the alternation. Finally, this analysis has suggested the need for a verb classification list for Latin, such as that of Levin 1993 for English, and in addition language-specific tests for eventuality structure in Latin are also required. Once these are accomplished, then the assessment of the relationship of verbal semantics and PAC serialization will be more accurate.

11.3 Animacy

11.3.1 Introduction

Animacy is yet another potential constraint on influencing grammatical realization. It is uncontroversial that there exists a hierarchy of the type animate > inanimate, which language users exploit in building up linguistic structure (cf. i.a. Comrie 1989; Croft 2003; Siewierska 1988; Tomlin 1986). This is evidenced in many ways, as seen from grammatical asymmetries in corpus data and from differing comprehension and production rates in psycholinguistic and neurocognitive experimentation. For instance, there is an animate-first constraint that operates over linearization, and in general terms animate entities are more frequent, accessible and are processed and produced faster than inanimate ones.

11.3.2 Grammatical consequences of animacy

11.3.2.1 Categorical effects

In a number of languages animacy and its associated hierarchy is so pervasive that it has grammatically categorical consequences. I report briefly on some of these below, taken from the literature.

There are reflections of the distinction in predicate agreement. In Ancient Greek, NP plural subjects that are relatively low in animacy (i.e. neuter plurals) are inflected in the singular, whereas ones high in animacy (i.e. masculine and feminine plurals) are inflected in the plural (Comrie, 1989: 190–191). In many pronominal systems around the world, different pronouns are associated with different degrees of animacy. In Russian, for example, one finds *kto*, the relative pronoun for animates, whereas *čto* is used for inanimates (Comrie, 1989: 191). In Navaho (Eyak-Athabaskan), the verbal prefixes *bi-* and *yi-* can both be used when agent and patient are of the same animacy. However, when the agent is higher in animacy than the patient, then only *yi-* can be used (Hale 1972; Witherspoon 1980: 3; Siewierska 1988: 49; Comrie 1989: 193; Dahl & Fraurud 1996: 49–50; i.a.). In Jakalteko, a Mayan language, subject NPs of transitive predicates are constrained to be animate. A sentence with an inanimate subject, e.g. *[spesba cake te’ pulita]’ *the wind closed the door (lit. close wind the door)*,’ is ungrammatical (Siewierska 1988: 50; Dahl & Fraurud 1996: 49, with the example). This has also been asserted to be the case for Japanese and for Lakhota (Siouan) (for references, see de Swart et al. 2008: 132). In Welsh, intransitive impersonals can only occur if the interpreted subject is animate — you can’t say *disgynir* ‘it was fallen’ about a tree, but you could of a human (Arman, 2013). There are obligatory effects of animacy on word

---

12 The animacy hierarchy is part of Silverstein’s (1986) feature hierarchy.

13 Or at least they have been reported to be categorical; further probing might find some of the mentioned phenomena to be weakly probabilistic.

14 Ethnologue language names and spellings have been used throughout.
11.3. ANIMACY

order in ditransitives in Southern Sotho (Niger-Congo). Here, when the theme and the recipient are of equal animacy they can occur in either order. However, when they are of unequal animacy, the animate NP has to precede the inanimate NP (Siewierska, 1988: 56). There is also an effect of animacy on clitic order in Shambala (Niger-Congo), discussed briefly above with regard to number. Here non-human clitics precede human clitics, an apparent reversal of the animate-first hierarchy (Siewierska, 1988: 58). Thus there is a plethora of evidence that animacy has strong categorical effects cross-linguistically.

11.3.2.2 Probabilistic effects

In other languages, the animacy constraint operates probabilistically (cf. Dahl & Fraurud’s (1996: 50–51) “statistical tendencies”). The rich literature on the genitive alternation has found that animacy is the single most influential predictor of the choice between s-genitive and the of-genitive variants (Hinrichs & Szmrecsanyi, 2007; Rosenbach, 2003, 2005, 2002; Szmrecsanyi & Hinrichs, 2008; Szmrecsanyi, 2010; Tagliamonte & Jarmasz, 2008; Wolk et al., 2013). Although the operationalizations differ, these studies converge on the finding that animate possessors favour the s-genitive whereas inanimate possessors are more associated with the of-genitive.

Snider & Zaenen (2006) discuss the influence of animacy on fronted NPs in English. They find that, in the case of topicalized constructions, e.g. Beans I like, inanimate NPs are on the whole more likely to be topicalized.

Similar findings have been reported for the English dative alternation (Bresnan & Nikitina, 2003; Bresnan, 2007; Bresnan & Hay, 2008; Hay & Bresnan, 2006; Kendall et al., 2011; Wolk et al., 2013). In one of these studies, for instance, it is found that inanimate recipients favour the prepositional dative variant by 12.67 times compared to animate recipients (Bresnan et al., 2007).

Turning now to languages other than English, Dahl & Fraurud (1996) report a range of probabilistic effects of animacy for Swedish. In Swedish, transitive and intransitive subjects behave differently with respect to animacy, with the former more likely to have an animate subject NP ($\chi^2 = 833.2317, df = 1, p < 2.2e−16$). Similarly, animate referents are more frequently realized as indirect objects and with preposed NP modifiers compared than direct objects and with postposed NP modifiers, which are instead more likely to be associated with inanimate referents.

11.3.2.3 Animacy and AVC research

Despite the clear effects of animacy mentioned in the literature and its theoretical importance in linguistics and in psycholinguistic and neurolinguistic research, it is somewhat surprising to find that it has not been addressed as a potential factor on variable AVC serialization.

11.3.3 Variable profile and operationalization

I annotate for the animacy of the preverbal NP (or NP contained with a PP) (PreverbalAnimacy). Before I do this, I first discuss some operational difficulties.

As pointed out by Zaenen et al. (2004), several problems exist with animacy annotation. First, animacy is not a very well researched category (cf. Dahl & Fraurud 1996). Second it is very fluid, being sensitive not only to cross-linguistic differences but also to diachronic differences within the same language; thus, what is considered inanimate in one language may be considered animate in another, and what is inanimate in one time period may be differently animate in another time period (see e.g. Wolk et al. (2013) on how this affects the distribution of the dative and genitive alternations in the history of English). Third, while it is straightforward to identify, e.g., case forms, there is no intersubjectively robust way of determining animacy categories. Fourth, language is used creatively in figurative ways, particularly in idioms and fictional narrative, which can make animacy very difficult to annotate when NPs are so used. With regards to Latin, we are further faced with difficulties of (i) translating animacy
categories between languages (cf. problem two) and (ii) there are no native speakers to consult. This is very problematic, as very often what looks like an inanimate entity is treated as if it seems to be animate. For instance, an inanimate NP such as fortuna ‘fortune’ can often be picked up with animate anaphoric pronouns. Is fortuna therefore animate? On the other hand, Latin is not so problematic as English. For instance, while in English Rome can refer to the city as in Rome is a beautiful city, it can also be used to refer to the inhabitants of the city, e.g. Rome surrenders or Rome sent an embassy to Athens to sue for peace, in Latin only Romani ‘the people of Rome’ would be used in this instance.

Various annotation schemes for animacy have been preposed in the literature, from very simple ones such as animate vs. inanimate or human vs. non-human to more fine-grained ones, such as those of Zaenen et al. (2004), based around Garretson (2004), which involves twelve different levels. Here we will draw on the annotation scheme of Zaenen et al. (2004), coding for five levels human, organization, animal, concrete, and abstract. I go through these categories in turn now.

The level human is fairly self-explanatory, and codes for prototypical human beings as well as human-like referents, for instance mythological and anthropomorphic creatures. Examples of this from the present dataset include Caesar ‘Caesar’, mihi ‘to me’, and Eumemem ‘Eumenes’, and milites ‘soldiers’.

The level organization is operationalized slightly differently from that of Zaenen et al. (2004), who state “[t]he cut-off point between HUMAN and ORG was put at ‘having a collective voice/purpose’: so a group with collective voice and purpose is deemed to be an ORG, a group with collective action, such as a mob, is not an ORG”. To me, it is difficult to determine whether a group has a collective voice/purpose and there seems to be a great fluidity between the two senses, so both are included in the present scheme. As examples, we have hostem ‘enemy’, senatus ‘senate’, exercitus ‘army’, and cohortes ‘cohorts’.

For the level animal, it was decided to include one or more animals, as well as groups and collections of animals, for instance the NP pecu ‘flock’. Examples here include columbis ‘pigeons’, examen ‘swarm (of bees)’, reiculae ‘culled animals’, and porci ‘pigs’.

The level concrete is reserved for entities that are physically tangible, and might be paraphrased generically as ‘material’ or ‘matter’. These are in Garretson (2004)’s terms “prototypical concretes”. Examples for this level include litterae ‘letter’, scuta ‘shields’, litore ‘shore’, in aede Opism ‘in the temple of Ops’.

Abstract entities are non-concrete nominals, such as events nouns and such like, as in avaritiam ‘greed’, tarditas ‘slowness’, proelia ‘battle’, and co biennio ‘in that period of two years’.

11.3.4 Statistical modelling

The overall crosstabulation of this predictor and serialization choice is reported in Table 11.10. An impressionistic glance shows that there is no effect of animacy, which appears to be supported by an initial \( \chi^2 \)-test of independence \( \chi^2 = 5.1167, df = 4, p = 0.276 \). However, we will still proceed with statistical modelling, as there may be important effects to be found when the multilevel structure of the dataset is incorporated into the mix.

This variable uses only \( n = 1511 \) observations, meaning that we need to construct a fresh baseline model for it to assess the value of the predictor. Table 11.11 compares the \(-2LL\) (deviance) and \(AIC\) statistics for four models: a standard GLM fitted for just the intercept, a VCM containing varying lemma components, a VCM containing varying text components, and a VCM containing both varying lemma and text components. The table shows that while lemma effects reduce the deviance slightly, it is certainly not enough on the relevant difference in degrees of freedom \( 2 - 1 \) to be significant. Textual components, by contrast, are important: first, they result in a deviance reduction of 2059.237 − 1994.432 = 64.805, which is significant on a \( \chi^2\)-distribution with \( 2 - 1 \) degrees of freedom \( p < 0.001 \), and second the model

\(^{15}\)That is, I don’t make a distinction between animals and animal groups, comparable to humans and organizations (i.e. human groups).

\(^{16}\)Except for the fact that animal preverbal NPs disprefer factus est serializations more than on average, but one should take that with a pinch of salt, given the low frequency counts for this level.
11.3. ANIMACY

<table>
<thead>
<tr>
<th></th>
<th>Serialization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>est factus</em> (%)</td>
</tr>
<tr>
<td>abstract</td>
<td>250 (39.87)</td>
</tr>
<tr>
<td>concrete</td>
<td>181 (46.06)</td>
</tr>
<tr>
<td>human</td>
<td>165 (43.08)</td>
</tr>
<tr>
<td>organization</td>
<td>36 (38.71)</td>
</tr>
<tr>
<td>animal</td>
<td>8 (53.33)</td>
</tr>
<tr>
<td>Totals</td>
<td>640</td>
</tr>
</tbody>
</table>

Table 11.10: Relationship of preverbal animacy and PAC serialization

<table>
<thead>
<tr>
<th>Model</th>
<th>$-2LL$</th>
<th>$AIC$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM null</td>
<td>2059.237</td>
<td>2061.237</td>
</tr>
<tr>
<td>VCM + varying lemma components</td>
<td>2058.119</td>
<td>2062.119</td>
</tr>
<tr>
<td>VCM + varying text components</td>
<td>1994.432</td>
<td>1998.432</td>
</tr>
<tr>
<td>VCM + varying lemma and text components</td>
<td>1993.519</td>
<td>1999.519</td>
</tr>
</tbody>
</table>

Table 11.11: Constructing a baseline VCM for animacy effects

containing them has the lowest $AIC$. We should therefore select this model as the baseline VCM of interest.

To turn to statistical modelling now, two multilevel models are built, with the first simply specifying the predictor term and varying intercepts for text:\footnote{Recall that as per Table 11.11 we do not need to include varying intercepts for lemma; there may be different differences with respect to the slopes though, but it is usually not the done thing to fit a model with varying slopes, but without including varying intercepts as well. In any case, a model so fitted with the additional term $(0+\text{Animacy}|\text{Lemma})$ results in the highest $AIC$ of 2029.19.}

\[
\text{anim.ri}\leftarrow \text{glmer}(\text{Serialization}\sim\text{Animacy}+\text{Text}),\text{anim, family=binomial})
\]

The second model contains varying slopes for text:

\[
\text{anim.rs.text}\leftarrow \text{glmer}(\text{Serialization}\sim\text{Animacy}+\text{Animacy}|\text{Text}),\text{anim, family=binomial})
\]

The ANOVA output in Table 11.12 demonstrates that the inclusion of varying slopes for text does not significantly improve the model. We can do without them, and proceed to compare \texttt{anim.ri} with the baseline VCM in Table 11.13. This shows that the addition of the animacy variable is superfluous, and the basic model should be preferred. There is no value in considering animacy a predictor of PAC serialization.

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chi sq</th>
<th>Chi DF</th>
<th>Pr(&gt;Chi sq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>anim.ri</td>
<td>6</td>
<td>2005.59</td>
<td>2037.52</td>
<td>-996.80</td>
<td>1993.59</td>
<td>8.60</td>
<td>14</td>
<td>0.8557</td>
</tr>
<tr>
<td>anim.rs.text</td>
<td>20</td>
<td>2024.99</td>
<td>2131.40</td>
<td>-992.50</td>
<td>1984.99</td>
<td>8.60</td>
<td>14</td>
<td>0.8557</td>
</tr>
</tbody>
</table>

Table 11.12: ANOVA statistics for various glmer$s containing animacy information

11.3.5 Discussion

It is striking that animacy appears to play absolutely no role in PAC serialization choice, despite the fact that it is otherwise important grammatically in the (non AVC) case studies reported above. It might be argued, however, that one possible reason that animacy is not influential is that it is not entirely independent of concreteness (see Gries 2003 for the irrelevance and epiphenomenality of animacy for the English particle alternation when pitted against concreteness). This possibility is to be rejected. If the levels concrete, human, and animal are contrasted with abstract and organizations (which are abstracts of sorts), there is no support for a concreteness model either (ANOVA: $\chi^2 = 0.3822$, df = 1, $p = 0.5364$).
### Table 11.13: ANOVA of baseline VCM and selected GLMER containing animacy information

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.anim</td>
<td>2</td>
<td>1998.43</td>
<td>2009.07</td>
<td>-997.22</td>
<td>1994.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anim.ri</td>
<td>6</td>
<td>2005.59</td>
<td>2037.52</td>
<td>-996.80</td>
<td>1993.59</td>
<td>0.84</td>
<td>4</td>
<td>0.9330</td>
</tr>
</tbody>
</table>

It therefore appears that previous researchers examining AVC/PAC alternations in Latin, historical German, and Belgian Dutch are therefore exonerated from including it as a potential source of variation, at least as a main effect. Even so, it could be that animacy interacts with other influential predictors, a possibility that will be explored later in the thesis.

### 11.4 Information structure

We move on in this section to motivate the inclusion of variables concerning information structure within the information space of the PAC alternation. By the term *information structure* I mean how information is transmitted between the speaker and the hearer in order to update the hearer’s store of knowledge about the universe of discourse (the “common ground”) and how that information is linguistically expressed so as to maximize processing and parsing efficiency. It has to do with various distinctions, such as *old-new*, *topic-comment*, *focus-background*, and, where it is different from focus, the notion of *contrastiveness*.

#### 11.4.1 Dimensions of Information Structure

In this section I familiarize the reader with the various dimensions of information structure mentioned above in section 11.4, both to facilitate a reading of the next section and an understanding of the operationalization efforts.

The *old-new* dimension has been variously termed *information status* (Prince, 1992; Birner & Ward, 1998; Nissim et al., 2004; Calhoun et al., 2005), *familiarity status* (Prince, 1991), *cognitive states* (of identifiability/activation) (Lambrecht, 1994), and *givenness* (Schwarzschild, 1999; Krifka, 2007; Krifka & Musan, 2012, inter alia). I use the term *information status* here to refer to the extent to which a referent is available at a particular point in the discourse. For instance, in its most basic (i.e. *old-new*) conception, if I uttered *My dog barked*, I am making an assumption that my hearer can locate the referent of the expression *my dog*, whereas if I had uttered *A dog barked*, I would be making an assumption that my hearer cannot locate the referent of the expression *a dog*. However, once I have uttered *A dog barked*, the hearer can then locate the expression’s referent, and I can go on and refer to it by a definite pronoun or some other type of definite expression. We can refer to the referent of the expression *my dog* as being *old* at that particular point in the discourse when it is uttered, and to the referent of the expression *a dog* as being *new*. More complex articulations have been proposed, proceeding through the trichotomy *old-mediated-new* (Nissim et al., 2004; Calhoun et al., 2005) to the proposal of Lambrecht (1994: 109), who distinguishes seven information states, based on referents’ *identifiability* and (for identifiables) degree of *activation*. His scheme is presented in 11.1 for reference, along with traditional reference points for new and old.

Finally, it is worth pointing out that information status and *definiteness* are inextricably linked (Lambrecht, 1994; Lyons, 1999). Definiteness will play an essential role in the operationalization of information status here.

While information status concerns what Lambrecht (1994) terms *pragmatic states*, other dimensions of information structure — i.e. *focus-background* and *topic-comment* — concern what he terms *pragmatic...*  

---

18 This term was established by Halliday (1967), but apparently the notion underlying it can be traced back to Aristotle (Krifka & Musan, 2012: 1, 26). The term *information packaging* (Chafe, 1976) is sometimes used, but it is much less common than information structure. For instance, a Google Scholar search reveals 24,500 hits for the search term "information structure" linguistics, whereas that for "information packaging" linguistics reveals in comparison only 2,880 hits.
11.4. INFORMATION STRUCTURE

relations. With regard to the focus-background dimension, pragmatically and semantically, the “back-
ground” represents information that is shared between the speaker and the hearer (“common ground”),
while the “focus” updates the interlocutor’s knowledge store with information which is new relative to
the background, and hence a pragmatic relation (Lambrecht 1994; Erteschik-Shir 1997, 2007; Krifka 2007;
Krifka & Musan 2012, inter alia). Take the sentence John had lamb chops for supper. If it is known that
John had something for supper, then John had ___ for supper is the background (sometimes called the
“open proposition”), and the np lamb chops is the focus of the assertion. Thus, very basically, the focus
corresponds to the answer in a wh-question. The scope of focus may concern the sentence (28a), the vp
(28b), an argument or adjunct constituent (28c), or a polarity operator (28d):

(28)    a. A: What happened?
        B: [A plane crashed]_{F}

    b. A: What happened to the plane?
        B: The plane [crashed]_{F}

    c. A: When did the plane crash?
        B: The plane crashed [last Tuesday]_{F}

    d. A: Did the plane crash on Tuesday?
        B: Yes, the plane [did]_{F} crash on Tuesday

Following Lambrecht (1994), examples (28a) and examples (28b) are instances of broad scope focus,
whereas examples (28c) and (28d) are instances of narrow focus, with the latter usually termed polarity
or verum focus as it has an effect on the truth value of the proposition that is being asserted (Höhle,

The second pragmatic relation is that of the topic-comment articulation (also termed theme-rheme in
the functional sentence perspective approach, e.g. Panhuis 1982; Firbas 1992). In very basic terms, the
“topic” is what the sentence is about, and the “comment” is what is said about the topic. For instance
in (29), my rabbit serves as the topic and is three years old, and really cute is what is said about it.

(29)    A: Tell me about your pets.
B: I’ve only got one pet. [My rabbit]_T is three years old, and really cute.

Example (29) illustrates the notion of sentence topics, but there can also be topics at the discourse level. In such a case, we may talk about the topic as being what the discourse is about, and the comment is all the information in (e.g.) a paragraph or a chunk of text that is said about the discourse topic. An example of this is in (30), where Barnaby can be considered the discourse topic, as the chunk of text is about him.\(^{19}\)

(30) My rabbit, he’s called Barnaby. My cat always likes to chase him around the house, but not in any kind of malicious way. My dog died last April, and he got on well with Barnaby too.

Another facet of information structure that may be identified is that of contrastiveness, which in some approaches is lumped together with focus. The view of focus proposed by Rooth (1992) is that it implies members of a set. For instance, if I say Mary likes football it evokes the set of sporting activities \{rugby, tennis, swimming, \ldots\}. However, topics can also be contrastive, as in the following example taken from (Krifka & Musan, 2012: 30).

(31) A: What do your siblings do?
B: [My sister]_T studies medicine, and [my brother]_T is working on a freight ship.

Some scholars talk of focus being overlaid onto the topic (“contrastive topic”) (e.g. Devine & Stephens 2006; Erteschik-Shir 2007). One can think of the difference between focus and contrast, as follows: (i) while focus may imply a set with which it is being implicitly contrasted, there is no intention of drawing or evoking the contrast set; (ii) contrast, on the other hand, is explicitly speaker intentional: the speaker intends to draw a contrast or a comparison between members of the same set.

In summary, I have here identified several dimensions of information structure — namely, information status, focus-background, topic-comment, contrastiveness. In the next sections, we will see why and how they are relevant for linguistic structure and word order alternation.

11.4.2 Linguistic and grammatical effects

11.4.2.1 General linguistic importance

The importance of these informational distinctions are widely recognised throughout the sub-disciplines of linguistics. For instance, they are influential in phonetics, where foci tend to receive pitch accents, are durationally longer relative to their syllable length, and are perceptually clearer, while backgrounded elements typically show the opposite tendencies. In morphology, foci and backgrounded elements are marked differently in some languages: for example, in Latin, foci can co-occur with focus-sensitive particles, such as quidem, whereas backgrounded elements cannot — puer quidem bonus est means ‘the boy is good’ or ‘the person who is good is the boy’, whereas puer bonus quidem est would convey a different interpretation ‘the boy is GOOD’ or ‘the quality that the boy has is that he is good’; certain languages of Africa are well known for the morphological marking of focus (see the contributions in Aboh et al. 2007). The distinction is also of importance in syntax, particularly in discourse-configurational languages (Kiss 1995, 1998, 2001): it has been argued that in such languages there are dedicated structural nodes, such as cross-categorial FocXPs, for focal entities that are situated above the xps which they c-command. Finally, the differences are in evidence in psycholinguistic and neurolinguistic research: in contrast to backgrounded elements, foci show longer fixation durations in eye-tracking studies, result in a larger N400, and induce positive ERP deflections, known as the ‘focus positivity’ effect (see e.g. Bornkessel-Schlesewsky & Schlesewsky 2009: 247ff. for discussion).

\(^{19}\)For a pragmatically-oriented view of the distinction between sentence and discourse topics, see e.g. Van Dijk (1977).
11.4.2.2 Effects in AVCs

The potential effect of information structure on internal AVC serialization has been noted, but its cross-linguistic relevance is up for debate. One the one hand, De Sutter (2009: 227), researching Dutch AVCs, notes that pragmatic effects of information structure are irrelevant to serialization differences, claiming that “pragmatic topicality measures [cannot] be applied to intraconstituent [i.e. AVC-type] syntactic variation. Moreover, one of the alternating elements, viz. the auxiliary verb, is semantically empty (apart from its aspectual meaning), so that the alternation cannot have any semantic or pragmatic consequences”. On the other hand, for historical German (Sapp, 2011) and for Latin (Walden, 1896; Marouzeau, 1909, 1910, 1938, 1953; Muldowney, 1937; Vogel, 1937; Wilkins, 1940; Adams, 1994; Devine & Stephens, 2006; Brookes, 2009; Spevak, 2010), information structure has been argued to be influential. Let us now discuss the purported effects in more detail.

In historical German, Sapp (2011: 24–29) observes an effect of information structure on verb cluster word order in Middle High German. He observes that, sentences that contain a constituent bearing either focus or contrast weakly favour the “1-2” order (i.e. auxiliary-nonfinite order) at a rate of 31%, compared to a baseline average of 28.7% for “1-2”, while sentences that do not exhibit an evident focus or contrast strongly prefer the “2-1” order (i.e. the nonfinite-auxiliary order) at a rate of 84.5%, compared to a baseline average of 71.3%. He shows that the type of constituent within the clause that bears the focus can also have an influence on the word order. For instance, focussed preverbal objects more readily give rise to the “1-2” order than the baseline (50.7% vs. 28.7%).

Turning to Latin, it has been argued that the factus est variant is used when the lexical verb (the participle) is “the main information asserted in the nuclear clause”, i.e. when it is the focus of the clause, whether broad or narrow scope, or in some way the only contrastive element in the clause (Devine & Stephens, 2006: 180; see also Spevak, 2010). I now illustrate these points with some fairly clear-cut examples, (32)–(34), taken from the previous literature. Example (32) is an announcement about the birth of a child, thus, arguably, focusing on the entire clause, and answering ‘What has happened?’ and an example of broad scope focus. In (33), the referent of the preverbal constituent Orgetorix, someone’s name, is available information in the preceding discourse; what is new information is what happened to him — namely that he died. His death is then topicalized in the subsequent clause in mortem. Thus, mortus est is arguably the focus of this sentence, and could plausibly answer ‘What happened to Orgetorix?’. In example (34) the participle is obviously contrastive, where the manner of the motion is being contrasted.

(32) natus est nobis nepos
born.PTCP.PL.M.SG.NOM be.3SG.PRS.IND US.M.PL.DAT grandson.M.SG.NOM

‘A grandson has been born to us’ (Ter. Hec. 639)

(33) Orgetorix mortuus est ...
Orgetorix.M.SG.NOM ... Orgetorix.M.SG.NOM die.PTCP.PL.M.SG.NOM be.3SG.PRS.IND ...
post eius mortem
after PRN.M.SG.GEN death.F.SG.ACC

‘Orgetorix ... Orgetorix died... after his death’ (Caes. Gal. 1. 5)

(34) non profectus est sed profugit
NEG depart.PTCP.PL.M.SG.NOM be.3SG.PRS.IND but flec.3SG.PRF.IND

‘he did not depart, he fled’ (Cic. Phil. 5. 24)

By contrast, the est factus variant has been argued to be used when the eventuality encoded by the verb is presupposed or easily accommodated information, and the preverbal constituent is under narrow focus or is contrastive (Devine & Stephens, 2006; Spevak, 2010). Thus, Devine & Stephens observe: “If on
the other hand the main information being communicated is not the occurrence of the event encoded by the verb but the identity of one of the participants in the event or of some circumstantial factor, then the auxiliary raises to the head of the focus projection and consequently appears to the left of the participle” (2006: 181). For example, “Contrast facile sunt repulsi [lit. ‘they were easily defeated’] . . . with facile experti sunt [lit. ‘they easily learned’] . . .: in the former the adverb carries the focus (‘the defeat of the Pompeians was an easy matter’), in the latter it is an adjunct to the verb phrase focus (‘they soon learned’)” (2006: 182). To illustrate this in operation in a discourse setting, consider (35). Here, what is under question is where various parts of the army have been stationed. Thus, the PAC erant conlocatae ‘were/had been stationed’ is discoursally available (cf. in eo loco ‘in that place’) and in dextro cornu ‘on the right wing’ is contrastive with the other positions that parts of the army could occupy (e.g. medium aciem).

(35) in eo loco ipse erat Pompeius. in PRN.M.SG.ABL place.M.SG.ABL self.M.SG.NOM be.3SG.IMP.IND Pompeius.M.SG.NOM. medium aciem Scipio cum legionibus middle.of.F.SG.ACC battle.line.F.SG.ACC Scipio.M.SG.NOM with legion.F.PL.ABL Syriacis tenebat. Ciliciensis legio . . . in dextro Syrian.F.PL.ABL hold.3SG.IMP.IND. Cilician.F.SG.NOM legion.F.SG.NOM . . . on right.N.SG.ABL cornu erant conlocatae wing.N.SG.ABL be.3PL.IMP.IND place.PTCP.PL.PRF.F.PL.NOM ‘In that place was Pompeius himself. Scipio held the middle part of the battle-line along with the legions from Syria. The legion from Cilicia had been stationed on the right wing.’ (Caes. Civ. 3. 88. 3)

Another informational factor that has been claimed to motivate the est factus serialization is polarity focus, where “what is at issue is the truth or falsity of the whole proposition” (Devine & Stephens, 2006: 165) or when there is a contrast of, e.g., tense or mood. As an example of this pattern, consider (36), taken from Adams (1994), who also mentions the same polarity focus effect. Here, the positive polarity of the second clause is explicitly contrasted with the negative polarity of the first clause.

(36) qui aut non est uictus umquam ... aut si REL.M.SG.NOM or NEG be.3SG.PRS.IND defeat.PTCP.PL.PRF.M.SG.NOM ever ... or if est uictus ... be.3SG.PRS.IND defeat.PTCP.PL.PRF.M.SG.NOM . . . ‘either he was never defeated ... or if he was defeated ...’ (Cic. De orat. 3. 129)

11.4.3 Variable profile and operationalization

11.4.3.1 Problems of operationalization

In the previous sections we have seen that information structure is an important linguistic concept with several dimensions that have effects on linguistic structure. However, one of the main problems with information structure is that it is difficult to operationalize in an empirically reliable way, which as noted in Chapter 7 is essential. Let us discuss how previous operationalizations have been problematic both for AVC/PAC alternations specifically and in general terms.

Consider the study of Spevak (2010: 39). For Spevak, “focus” means “the most important or salient piece of information in a clause”. A number of “criteria” are established to distinguish a focal entity from a non-focal entity. For instance, Spevak states “[there] are a few formal indications of salience: enumeration, repetition, and coordination” (2010: 39). I call this Criterion 1. Another criterion that is mentioned is that of the “question test”: “[the] most informative part of the sentence can be determined

20 Though, he calls it “veridical focus”.

220
with the help of a question test. Indeed, each sentence can be viewed as the reply to an explicit or implicit question” (2010: 35). I call this Criterion 2. What are the problems with such criteria? In the following example, (37), Spevak regards the verb (interficio ‘kill’) as being focal.

(37) quibus submotis omnes sagittarii funditores=que
    REL.M.PL.ABL remove.M.PL.ABL all.M.PL.NOM archer.M.PL.NOM slinger.M.PL.NOM=and
    destituti inermites sine praeidio
destitute.M.PL.NOM unarmed.M.PL.NOM without protection.N.SG.ABL
    interficti sunt.
    kill.PTCP.PFR.M.PL.NOM be.3PL.PRS.IND

‘When they had gone, all the archers and slingers, left unarmed and defenceless, were killed’
(Caes. Civ. 3. 93. 7, in Spevak 2010, 154, ex. (21))

However, according to Criterion 1, it is difficult to regard to the verb as the focus: destituti inermites sine praeidio is an asyndetically coordinated adjunct phrase, which repeats virtually the same idea. Going on Spevak’s criteria, one might plausibly identify this adjunct phrase with focus. What about Criterion 2? In (38) below, Spevak says that profectus est tells us the answer to ‘what does Domitius do?’. It is thus, apparently, the focus.

(38) sed paucis ante diebus L. Domitius . . . unam
    but few.M.F.PL.ABL before day.M.F.PL.ABL L. Domitius.M.SG.NOM . . . one.F.SG.ACC
    [nauem] ipse conscenderat . . . , nancetur
    ship.F.SG.ACC self.M.SG.NOM embark.3SG.PL.PRF.IND . . . , obtain.PTCP.PRF.M.SG.NOM
    turbidam tempestatem profectus est.
    wild.F.SG.ACC weather.F.SG.ACC depart.3SG.PL.PRF.M.SG.NOM be.3SG.PRS.IND

‘However, a few days previously Lucius Domitius . . . embarked on the third ship himself, and took advantage of some wild weather to set out’ (tr. Spevak)
(Caes. Civ. 2. 22. 2, in Spevak 2010, 156, ex. (28))

However, one could equally regard the implicit question here as the following: ‘so Lucius Domitius got on the ship, when did he set out?’ with the focus now on the adjunct phrase nancetur turbidam tempestatem. The idea profectus est is inferred from the notion of getting on the ship: why else would a person get on a ship?

Devine & Stephens (2006) and Marouzeau (1910) similarly offer no empirically robust diagnostics. For instance, Marouzeau (1910: 113) asserts “factus est sert à définir l’action . . . est factus sert à affirmer ou à confirmer la réalité de l’action”, but no attempt is made to offer an operational definition, and thus his claims fail to become testable. Similarly, Sapp (2011: 62), working on historical German AVCS, vaguely says on tagging contrastiveness: “Clauses were tagged contrastive regardless of whether an argument was new or given in the discourse, if it could be interpreted as contrastive”. But it is not stated how to identify contrastiveness. Consequently it is very difficult for us to verify their claims as it is impossible to know exactly what the original authors meant.

Thus far we have only offered anecdotal evidence for the problem of coding information structure according to weak operational definitions. However, Cook & Bildhauer (2011) discuss the problem of the notion “aboutness topic” in detail, providing statistical evidence that it is indeed very problematic. In their study, the following criteria, based on Götze et al. (2007: 165), are used to identify the aboutness topic of a sentence.

(39) An NP X is the Aboutness Topic of a sentence S containing X if
    a. S would be a natural continuation to the announcement
       Let me tell you something about X
    b. S would be a good answer to the question
       What about X?
c. S could be naturally transformed into the sentence

Concerning X, S

where S differs from S only insofar as X has been replaced by a suitable pronoun.

Two annotators follow the above guidelines and tag 587 sentences of German prose for the presence vs. absence of aboutness topics. Cook & Bildhauer (2011) use Cohen’s $\kappa$ statistic to assess the inter-annotator agreement. Overall, they find that $\kappa$ does not exceed 0.57, “which in our view is much less than could be expected in a case where both annotators base their judgements on the same guidelines”. This finding is firm confirmation of the problems faced when using untested and unreliable diagnostics to code datasets.

Researchers are acutely aware that the enterprise of annotating for information structure is subjective. For example, Devine & Stephens (2006) say that “[w]hile we have tried to cite empirical evidence to support the informational structures just analyzed, our account sometimes involved relatively subjective judgements about speaker intentions and perspective, which could lead to some circularity and which are harder to corroborate in a dead language like Latin”. In a similar vein, Sapp (2011: 62) says “[u]sing context for determining the focus structure of a sentence in a non-living language can be quite difficult, and that is especially true for contrastive focus. It is possible that in some cases the 1-2 order has biased me into reading the clauses with a special intonation, and there are no doubt instances of focus that I have overlooked”. But I believe that if one wants to annotate for information structure, it must be done only in a way that is cross-study and cross-annotator reliable. Against this backdrop, I therefore introduce in the following sections several “quick and dirty” ways to measure information structural concepts, and provide detailed coding guides where that is not possible.

11.4.3.2 Delimitation of the Discourse Context

Some of the variables and their operationalizations that follow make use of the notion of discourse context. The definition here makes use of the LLT-A text archive from which the observations were extracted. I define it as the preceding ten lines of text and the following ten lines of text.

A summary of the discourse context for one particular data point is given in Figure 11.2.

<table>
<thead>
<tr>
<th>Line Distance</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>−10</td>
<td>regna conciliaturum confirmat.</td>
</tr>
<tr>
<td>−9</td>
<td>hac oratone adducti inter se fidem et ius iurandum dant et regno occupato per tres potentissimos ac firmissimos populos totius Galliae sese potiri posse sperant.</td>
</tr>
<tr>
<td>−7</td>
<td>Ea res est Helvetii per indicium enuntiata.</td>
</tr>
<tr>
<td>−6</td>
<td>moribus suis Orgetorigem ex vinculis causam dicere coegerunt; damnum poenam sequi oportebat ut igni cremaretur.</td>
</tr>
<tr>
<td>−5</td>
<td>die constituta causae dictionis Orgetorix ad judicium omnem suam familiam, ad hominum milia decem, undisque coegit et omnesclientes obaeratos que suos, quorum magnum numerum habebat, eodem conduxit; per eos, ne causam dicere, se eripuit.</td>
</tr>
<tr>
<td>−1</td>
<td>cum civitas ob eam rem incitata armis ius suum exsequi conaretur multitudinem que hominum ex agris magistratus cogerent, Orgetorix mortuus est; neque abest suspicio, ut Helvetii arbitrantur, quin ipse sibi mortem consciverit.</td>
</tr>
<tr>
<td>2</td>
<td>Post eius mortem nihilominus Helvetii id, quod constituerant, facere conantur, ut e finibus suis exerant.</td>
</tr>
<tr>
<td>3</td>
<td>ubi iam se ad eam rem paratos esse arbitrati sunt, oppida sua omnia numero ad duodecim, vicos ad quadririgentes, reliqua privata aedificia incendunt, frumentum omne, praeter quod se cum portaturi erant, comburunt, ut domum reditionis spe sublata paratiore ad omnia pericula subeunda essent; trium mensum molita cibaria sibi quemque domo effere iubent.</td>
</tr>
<tr>
<td>4</td>
<td>persuasent Rauracis et Tulingis et Latobicis finitimis uti eodem usi consilio oppidis coniventum multitudinem que hominum ex agris una cum iis profisciscantur, Boios que, qui trans Rhenum incoluerant et in agrum Noricum transierant Noreiam que oppugnarant, receptos ad se socios sibi adsiciscunt.</td>
</tr>
<tr>
<td>5</td>
<td>Erant omnino itinera duo, quibus itineribus domo exire iubent.</td>
</tr>
</tbody>
</table>

Figure 11.2: Discourse context for mortuus est (Caes. Gal. 1. 4. 3). ‘Line distance’ represents the backward and forward line distance from the line containing the PAC.

21 Of course, language users make use of more than just the wide discourse context, thus defined, such as situational or world knowledge, as in Lambrecht (1994), but this is much more difficult to pin down.
11.4. INFORMATION STRUCTURE

11.4.3.3 Definiteness

It is of interest to determine whether the information status of the preverbal constituent has an effect on PAC serialization. To approximate information status it was decided to annotate for the definiteness of the preverbal NP or PP, because the two are clearly related (indefinite NPs are necessarily discourse new) and it is less time consuming to tag. While definiteness is not grammatically marked in Latin, it is typically marked in the English translation by means of indefinite and definite articles, for instance. As a further test of definiteness, one can apply the existential-there diagnostic mentioned by Garretson (2004):

(40) There is/are NP outside.

If the NP fits felicitously into the gap it can be considered indefinite; if this is not the case, it is considered definite. To exemplify, in (41), the preverbal constituent is indefinite because it felicitously fills the slot after there are.

(41) castra [opportunis locis] erant posita

‘Camps had been situated in opportune locations’ (Caes. Gal. 7. 69. 1)

Garretson (2004) provide a helpful list of all English word forms that necessarily contribute to either an indefinite or definite reading of the NP. This can also be used to help identify the definiteness of the preverbal expression. This predictor is termed PreverbalDefiniteness and comprises two levels definite and indefinite. Note that only NP/PPs are included here; all other phrasal and word categories are tagged as missing datapoints. All in all, there are \(n = 1178\) datapoints with the level definite, \(n = 333\) with the level indefinite. There are \(n = 898\) missing datapoints for this variable.

11.4.3.4 Topicality

Another salient indicator of focus is whether an expression is picked up (or “topicalized”) in what follows in the subsequent discourse. An NP is considered to be picked up, if it forms a reference chain, whereby it is referred to in the subsequent clause by (i) a relative pronoun, (ii) an anaphoric or other demonstrative pronoun, (iii) lexical repetition, or (iv) becomes the null subject of the subsequent verb. All NPs are included in this annotation, with unmarkable non-NP phrase types (e.g. clauses, conjunctions, adverbs, white space, etc.) tagged as missing datapoints. To exemplify, in (42), the preverbal NP opportunis locis ‘opportune locations’ is picked up by the demonstrative adverb ibi ‘there’ in the following clause.

(42) castra [opportunis locis] erant posita ibique castella XXIII facta.

‘Camps were/had been situated in opportune locations and twenty three fortresses had been constructed there’ (Caes. Gal. 7. 69. 1)

The variable is termed PreverbalTopicality and takes two levels: yes if the NP is picked up and no if it is not.

11.4.3.5 Contrastiveness

Contrastiveness is difficult to operationalize because of its inherent subjectivity. Nonetheless, it is still important to include in the PAC’s information space. In order to do so and circumvent the difficulties and subjectivities involved, I will extensively outline those grammatical, lexical and pragmatic contexts in which it can be supposed that a contrastive (or focal) interpretation is possible. Others may disagree with the contexts that I have chosen to include as being contrastive, but I submit that doing it this way provides a well-defined operationalization of a difficult variable and allows for replicability.

11.4.3.5.1 Grammatical contexts
11.4.3.5.1.1 From-to In these examples there is a from-to contrastive structure involving frames such as from place to place, from person to person, and from time to time, etc.

(43) profectus ab Utica in Mauretaniam regnumque Bogudis est ingressus
    ‘Having departed from Utica he entered Mauretania and the kingdom of Bogus’ (B. Afr. 23)

(44) ab Iuba ad eum est missus
    ‘He was sent from Juba to him’ (B. Afr. 57)

(45) ab hora diei quinta usque ad solis occasum est decertatum
    ‘Fighting took place from the fifth hour of the day right up until sunset’ (B. Afr. 19)

11.4.3.5.1.2 Reciprocal This type can be illustrated by the English John hit Bill and Bill hit John, where the second clause is effectively a consequence of the first clause, and the referents switch argument roles (Ladd, 2008; Rooth, 1992). For instance, in this example, John is the agent and Bill is the patient in the first clause, but in the second clause the argument roles switch: Bill is now the agent and John is now the patient. Another example would be You scratch my back, I’ll scratch yours. The contrast here is the argument switch.

(46) hic cum multa regi esset pollicitus . . . namque hanc urbem ei rex donarat
    ‘Since he had promised the king many things . . . for the king had bestowed this city upon him’
    (Nep. Them. 10. 2)

(47) quem Dion . . . admiratus est . . . neque uero minus ipse Plato delectatus est Dione.
    ‘Dion admired him . . . and Plato himself was no less delighted by Dion’ (Nep. Dion 2. 3)

11.4.3.5.1.3 Juxtaposition The juxtaposed type involves the syntactic juxtaposing of two identical lexical items which have distinct referents. The contrast here involves the reference switch involving the same lexical item.

(48) Quo saluo salui sumus futuri.
    ‘If this is safe, we will be safe’ (Cic. Phil. 11. 24)

(49) cum castra castris conlata essent
    ‘Since camp was joined to camp’ (B. Alex. 61.2)

11.4.3.5.1.4 Not-but This type is a prototypical contrastive structure involving not X but Y or Y but not X, where the mentioning of Y serves to exclude (or replace) X. They are thus replacive contrasts, and are straightforward to identify. In the straightforward case, the verb is the same or has the same semantic construal (perhaps even ellipted in one member of the contrast), and there is a single contrast between the two arguments or adjuncts, one of which appears left-adjacent to the pac. The not X but Y contrast may induce a double contrast, where both arguments/adjuncts and their respective verbs are contrastive with each other.

(50) Antoni est acta causa ab amicis eius, non publica.
    ‘It is Antonius’ cause that has been pleaded by his friends, not the state’s’ (Cic. Phil. 12. 3)

(51) mora est adlata bello, non causa sublata.
    ‘Delay has been brought to the war, the reason for it has not been removed’ (Cic. Phil. 6. 1)

11.4.3.5.1.5 Either-or Another structure that induces a contrastive interpretation involves either X or Y (or in Latin aut X aut Y). They are instances of exclusive disjunction, where the selection of X automatically excludes the selection of Y, and vice versa.
11.4. INFORMATION STRUCTURE

11.4.3.5.1.6 Not-only-but-also These structures are similar to the not X but Y structure of replacive focus, but here it is additive and cumulative rather than replacive, in that the addition of the Y term “amplifies the set of focus entities by adding one (y) that he thinks is not entertained by the listener” (Devine & Stephens, 2006: 226).

11.4.3.5.1.7 Both-and This is another type of additive structure, similar to not only X but also Y.

11.4.3.5.1.8 Clarificatory In these structures, there are two juxtaposed terms X and Y, where X and Y are equative in that they share the same referent and Y serves to clarify (or expand upon) the referent of X, or vice versa. For example, in (54) hoc (Y) ‘this’ refers back to the adjacent previous expression liberatis suis (X) ‘with his own men having been released’; here, X clarifies Y. In (55) stramentis (Y) ‘with thatch’ refers back to more Gallico (X) ‘in the Gallic fashion’; here, X is clarified by Y.

11.4.3.5.1.9 Comparative Comparatives are treated as being explicitly contrastive when they introduce a following quam-clause (than-clause), but not otherwise. Here the preverbal term is linked to an expression in a following or preceding comparative clause that it is contrasted with.

11.4.3.5.1.10 Predicative A further grammatical structure in which focus (or contrast) seems to be evoked is in predicative complement structures, particularly with predicates such as call and name. For instance, it is fairly predictable that someone is called something.

11.4.3.5.2 Lexical contexts In these examples, there is a local in-sentence lexical contrast, where the two contrasting elements are different arguments (or adjuncts) of the verb. Similar to from-to contrasts, these are a type of symmetric contrast, identified by Rooth (1992). In order to qualify, the lexical items under contrast must be antonyms, which can be checked on WordNet http://wordnetweb.princeton.edu/
via their English translation, or else they must belong to a plausible set in the local discourse context. Typical contrasts involve size and evaluation terms, or philosophical contrasts of death and being alive, mortality and immortality. In (59), *huius uitia ineuntis adolescenciae* ‘faults of his early youth’ are contrasted with *magnis ... uirtutibus* ‘great virtues’. In (60), there is a dead/alive contrast.

(59) huius uitia ineuntis adolescenciae magnis sunt emendata uirtutibus

‘The faults of his early youth were emended by his great virtues’ (Nep. Them. 1)

(60) uita enim mortuorum in memoria est posita uiiuorum.

‘For the life of the dead is placed in the memory of the living’ (Cic. Phil. 9. 9)

11.4.3.5.3 Pragmatic contexts

11.4.3.5.3.1 Polarity In these examples, there is an explicit contrast between positive and negative polarity. The arguments and the verbs are the same (or roughly the same in terms of their semantic construal), and one clause has positive polarity and the other clause has negative polarity.

(61) quae terra sint contentae . . . quae non sunt contentae terra

‘which are happy with the land . . . which are not happy with the land’ (Var. R.R. 3. 3)

11.4.3.5.3.2 Contrastive Expressions which are contrastive must share the same verb lemma, perhaps even being omitted in one member of the contrast, and must vary with respect to at least one argument/adjunct. Several examples will illustrate this. In (62), the king (*rex*) and Caesar are constructed as a set of individuals departing for Mithridates with different purposes in mind. Concerning the sentence containing the *pac*, the individuals in that set (the king, Caesar) are then contrasted with respect to the routes they use, namely the Nile and the sea, respectively. The verb is the same (cf. *uti noluit* ‘did not want to use’ in the next sentence). The preverbal constituent *rex* is therefore in antithesis with *Caesar* in the next sentence.

(62) ita paene sub idem tempus et rex ad opprimendum Mithridaten proficiscitur et Caesar ad recipiendum. celeriore fluminis Nili nauigatione rex est usus in quo magnam et paratam classem habebat. Caesar codem itinere uti noluit ne nauibus in flumine dimicaret sed circumuectus eo mari quod Africae partis esse dicitur sicuti supra demonstrauimus.

‘Accordingly at practically the same time the king set forth to crush Mithridates, and Caesar to relieve him. The king had recourse to the quicker method of transport, namely sailing up the river Nile, in which he had a large fleet in readiness. Caesar was unwilling to use the same route, so as not to fight a naval action in the river. Instead, he sailed round by that sea which is said to belong to part of Africa, as I have explained earlier.’ (B. Alex. 28. 1)

In (63) different parts of the army (*pars, pars*) are contrasted with respect to who killed them (*a militibus, ab equitibus*) and where they were killed (*ibi, portis*). Here, as in (66), the verb is omitted with the first member of the contrast.

(63) parsque ibi, cum angusto exitu portarum se ipsi premerent, a militibus, pars iam egressa portis ab equitibus est interfecta.

‘And part, as they crowded one another at the narrow passage of the gates, were slain there by the troops, part after they had got out of the gates by the cavalry’ (Caes. Gal. 7. 28)

In (64), there is a contrast between the types of buildings (*cella uinaria maior, ampliora . . . horrea*) that should be built on estates of different types (*uineta, frumentarius ager*). The verb is the same, being elided for the second member of the contrastive structure.

226
11.4. INFORMATION STRUCTURE

<table>
<thead>
<tr>
<th>ARGUMENT</th>
<th>VERB</th>
<th>ARGUMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>{king, Caesar}</td>
<td>use</td>
<td>{Nile, sea}</td>
</tr>
<tr>
<td>{soldiers, cavalry}</td>
<td>kill</td>
<td>{some, others}</td>
</tr>
<tr>
<td>{teeth, brows}</td>
<td>become</td>
<td>{white, grey}</td>
</tr>
<tr>
<td>{larger wine cellar, larger granaries}</td>
<td>be built on</td>
<td>{vineyard, grain farm}</td>
</tr>
</tbody>
</table>

Figure 11.3: Contrast matrix

(64) dubium enim non est quin cella uinaria maior sit facienda in co agro, ubi uineta sint, ampliora ut horrea, si frumentarius ager est. ‘There is no doubt that a larger wine cellar should be built on an estate where there is a vineyard, and larger granaries if it is a grain farm’ (Var. R.R 1. 11. 1)

In (65), different parts of the body (dentes, supercilia) are contrasted with respect to the different properties they begin to take on (the teeth become prominent, while the brows become grey). Again, the verb is elided in the second member of the contrast.

(65) hoc maiores qui sunt, intellegi negant posse, praeterquam cum dentes sint facti brocchi et supercilia cana et sub ea lacunae, ex observatu dicunt eum equo[n] habere annos sedecim. ‘It is said that there is no way of determining those which are older than this, except that when the teeth become prominent and the brows grey with hollows under them, they determine by looking at him that such a horse is sixteen years old (Var. R.R 2. 7. 3)

In (66), two qualities of an individual’s life (suam innocentiam, felicitatem) are contrasted with respect to when they have been witnessed (perpetua uita, Helvetiorum bello). The verb is the same, being elided this time for the first member of the contrast.

(66) suam innocentiam perpetua vita, felicitatem Helvetiorum bello esse perspectam. ‘My own blamelessness has been clearly seen throughout my life, my good fortune in the Helvetian campaign.’ (Caes. Gal. 1. 40)

As a “diagnostic” for this, we can enter the sentence into a contrast matrix in which the verb remains the constant term and at least one of the arguments and adjuncts are free to vary, as in Figure 11.3.

11.4.3.5.3.3 Parallel These are essentially double contrasts. While contrastive expressions necessarily involve the same verb (or one with similar semantics), these structures have different verbs associated with them. The similarity between this type and the former type is that the arguments/adjuncts of the verbs must form a plausible set, and be co-members/co-hyponyms of that set. The parallel type is particularly common in battle inventories, such as (67)–(68):

(67) ex Massiliensium classe V sunt depressae, IIII captae, una cum Nasidianis profugit. ‘from the Massilensian fleet five were sunk, four captured, one fled with the Nasidiani.’ (Caes. Civ. 2. 7)

(68) multi sunt interficti, complures capti ‘Many were killed, several were captured.’ (B. Hisp. 9. 3)

11.4.3.5.3.4 Superset Here, a unit is mentioned that comprises the general superset (hyperonym) for a preceding or upcoming contrast, where at least two members of the set are contrasted. The trigger can open or close the contrast. The general term is preverbal with respect to the pac. In (69), the gender is the grouped hypernym, and the mas and femina terms which follow are the hyponyms. In (70), the Athenians are being harassed from all sides; the narrator goes on to give examples of some of those
places. Hypernym and hyponym relations can be established or checked using WordNet via the English translations, if required.

(69) *mas an femina sit concepta*, significat descensu taurus, ... quod, si mas est, in dexteriorem partem abit; si femina, in sinistieriorem.

‘Whether it a male or female that has been conceived is signified by the way the bull dismounts, namely, if it is a male, it dismounts on the right, and if it is a female, it dismounts on the left’ (Var. R.R. 2.5)

(70) Athenienses undique premi *sunt coepti*: defecerat Samus, descierat Hellespontus, Philippus iam tum uales, Macedo, multa moliebatur.

‘The Athenians began to be pressed on all sides: Samos had defected, the Hellespont had revolted, Phillip of Macedon, already powerful, was stirring many things’ (Nep. Timoth. 3)

### 11.4.4 Statistical Modelling

#### 11.4.4.1 Definiteness

The overall relationship of the definiteness predictor and serialization choice is given in Table 11.14. There is an asymmetry whereby preverbal indefinite NPs disfavour *factus est* orders, an asymmetry which is statistically highly significant on a simple $\chi^2$-test of independence ($\chi^2 = 43.4079$, $df = 1$, $p < 0.001$).

<table>
<thead>
<tr>
<th>Serialization</th>
<th><em>est factus (%)</em></th>
<th><em>factus est (%)</em></th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>definite</td>
<td>446 (37.86)</td>
<td>732 (62.14)</td>
<td>1178</td>
</tr>
<tr>
<td>indefinite</td>
<td>194 (58.26)</td>
<td>139 (41.74)</td>
<td>333</td>
</tr>
<tr>
<td>Totals</td>
<td>640</td>
<td>871</td>
<td>1511</td>
</tr>
</tbody>
</table>

Table 11.14: Relationship of preverbal NP definiteness and PAC serialization

The baseline model is not the same as for the majority of variables as there are $n = 898$ missing datapoints. Therefore a baseline model specific to this variable needs to be established. The first model to be constructed is the basic single-level GLM. Three variance components models are then configured: one with variant components for lemma, a second with variant components for text, and a third with both. Table 11.15 shows the deviance ($-2LL$) and $AIC$ values for the several models considered. This demonstrates that out of the four models, the VCM with textual information is the most effective at reducing the deviance and providing a parsimonious fit at the same time. Variability according to lemma is simply not important for this subset of the data.

<table>
<thead>
<tr>
<th>Model</th>
<th>$-2LL$</th>
<th>$AIC$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM null</td>
<td>2059.237</td>
<td>2061.237</td>
</tr>
<tr>
<td>VCM + varying lemma intercepts</td>
<td>2058.238</td>
<td>2062.238</td>
</tr>
<tr>
<td>VCM + varying lemma and text intercepts</td>
<td>1993.767</td>
<td>1999.767</td>
</tr>
</tbody>
</table>

Table 11.15: Constructing a baseline VCM for definiteness effects

With the baseline VCM established, we are now in a position to consider the definiteness predictor. Two models are constructed this time. The first includes terms for the predictor and the varying intercept of text:

```r
definite.ri<-glmer(Serialization~Def+(1|Text),definite,family=binomial)
```

The second includes an additional term for varying slopes of text:

```r
definite.rs.text<-glmer(Serialization~Def+(Def|Text),definite,family=binomial)
```
As Table 11.16 demonstrates, adding the varying slopes of text does not significantly reduce the deviance of the model, and as such definite.ri is both structurally and informationally simpler.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>definite.ri</td>
<td>3</td>
<td>1961.02</td>
<td>1976.98</td>
<td>-977.51</td>
<td>1955.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>definite.rs.text</td>
<td>5</td>
<td>1964.91</td>
<td>1991.51</td>
<td>-977.45</td>
<td>1954.91</td>
<td>0.11</td>
<td>2</td>
<td>0.9460</td>
</tr>
</tbody>
</table>

Table 11.16: ANOVA statistics for various GLMERS containing definiteness information

This model is then compared with the baseline definiteness vcm constructed above. Table 11.17 indicates that the inclusion of the predictor substantially reduces the deviance from 1994.59 to 1955.02, a deviance of 39.57 which on a χ²-distribution with 3 − 2 = 1 degree of freedom is highly significant. Its parsimony index of 1961.02 as compared to 1998.59 of the baseline vcm indicates that it should clearly be preferred.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>definite.ri</td>
<td>3</td>
<td>1961.02</td>
<td>1976.98</td>
<td>-977.51</td>
<td>1955.02</td>
<td>39.57</td>
<td>1</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 11.17: ANOVA of baseline vcm and selected GLMER containing definiteness information

The parameters of the model are reported in Table 11.18. The logit probability is in terms of a factus est outcome, and the predictor variable is coded via dummy-coding such that the level definite is the baseline (coded 0). When the preverbal NP (or NP contained within a preverbal PP) is indefinite as opposed to definite, the logit probability of a factus est outcome decreases by −0.8124; this equates to an odds ratio of 0.44. The 95% confidence intervals associated with this parameter estimate do no cross zero, ranging from -1.067116287 to -0.5577704 in logits (or 0.3439991 to 0.5724841 in odds), and as a consequence it is highly significant. Note in addition that the absolute z-value for the slope is well above the critical value of 1.96 for large data samples. In sum, definite NPs stimulate factus est orders, while indefinite NPs stimulate est factus orders.

|                | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------|----------|------------|---------|---------|
| (Intercept)    | 0.3037   | 0.1590     | 1.910   | 0.0561  |
| Defndefinite   | -0.8124  | 0.1299     | -6.253  | 4.04e-10|

Table 11.18: Final GLMER for serialization ∼ preverbal definiteness

11.4.4.2 Topicality

Table 11.19 gives the overall distribution of preverbal topicality and serialization choice. As can be seen, there is a very slight asymmetry whereby preverbal NPs that are picked up in the subsequent clause are dispreferred with factus est.

The baseline vcm for this predictor is the same as that for definiteness, so we proceed directly to statistical modelling. Two models are configured, one with by-text random intercepts, the second with by-text random intercepts and random slopes:

topicality.ri<-glmer(RealizationOfPAC∼PreverbalTopicality+(1|Text),pac,family=binomial)
topicality.rs.text<-glmer(RealizationOfPAC∼PreverbalTopicality+(PreverbalTopicality|Text),pac,family=binomial)

The ANOVA output in Table 11.20 which compares the two models indicates that there is significant by-text variability in the slopes; not all texts respond the same with respect to their effect on the outcome variable (more on this below).
CHAPTER 11. SEMANTIC AND PRAGMATIC PREDICTORS

### Serialization

<table>
<thead>
<tr>
<th></th>
<th>est factus (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>465 (40.75)</td>
<td>676 (59.25)</td>
<td>1141</td>
</tr>
<tr>
<td>yes</td>
<td>175 (47.30)</td>
<td>(52.70)</td>
<td>370</td>
</tr>
<tr>
<td>Totals</td>
<td>640</td>
<td>871</td>
<td>1511</td>
</tr>
</tbody>
</table>

Table 11.19: Relationship of preverbal NP topicality and PAC serialization

<table>
<thead>
<tr>
<th>Model</th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>topicality.ri</td>
<td>3</td>
<td>1995.52</td>
<td>2011.48</td>
<td>-994.76</td>
<td>1989.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>topicality.rs.text</td>
<td>5</td>
<td>1992.15</td>
<td>2018.75</td>
<td>-991.07</td>
<td>1982.15</td>
<td>7.38</td>
<td>2</td>
<td>0.0250</td>
</tr>
</tbody>
</table>

Table 11.20: ANOVA statistics for various GLMERs containing topicality information

Before the selected model can be discussed further, we need to compare it against the baseline to determine whether the inclusion of the predictor is worth it or simply adds redundant information. The ANOVA in Table 11.21 indicates that the deviance is significantly decreased by adding the predictor term and its associated by-text slopes (1982.15 vs. 1994.59), and it has the lower AIC of the two models (1992.15 vs. 1998.59), despite the addition of three extra degrees of freedom.

### Anova of baseline VCM and selected GLMER containing topicality information

<table>
<thead>
<tr>
<th>Model</th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>topicality.rs.text</td>
<td>5</td>
<td>1992.15</td>
<td>2018.75</td>
<td>-991.07</td>
<td>1982.15</td>
<td>12.44</td>
<td>3</td>
<td>0.0060</td>
</tr>
</tbody>
</table>

An inspection of the model parameters is therefore warranted (Table 11.22). The level yes, indicating that the preverbal NP is picked up in the subsequent clause, is dummy-coded as 1, whereas the level no is established as the baseline. When the preverbal constituent is picked up as opposed to when it is not, the logit probability of a factus est outcome decreases by $-0.4362$, equivalent to an odds ratio of $0.6464884$. The absolute z-value is above the critical value of 1.96, such that we can be fairly confident that the population slope $\hat{\beta}$ would lie in negative logit space. The confidence intervals can be more specific about the range within which $\hat{\beta}$ might lie, namely from $-0.85344176$ to $-0.01903032$. However, the wideness of the confidence intervals suggests that there is much uncertainty about the specific value of $\hat{\beta}$.

Let us now evaluate the by-text random slopes for this predictor, which were seen above to enhance the model’s fit. These are plotted in Figure 11.4. The plot demonstrates the eight of the texts exhibit negative slopes, whereas the slope for Cicero’s Phillipics is positive, indicating his polar opposite behaviour.

#### 11.4.4.3 Contrastiveness

A crosstabulation of contrastiveness and serialization is presented in Table 11.23. As this indicates, preverbal expressions that are contrastive associate more with est factus serializations than factus est serializations, and by contrast preverbal expressions that are noncontrastive associate more with factus est serializations than est factus. The distribution is obviously statistically significant on a $\chi^2$-test of independence ($\chi^2 = 73.152$, df = 1, $p < 0.001$).

To turn to statistical modelling, four multilevel models are built, with the first simply specifying the predictor term and varying intercepts for lemma and text:

```r
costant.ri <- glmer(RealizationOfPAC ~ PreverbalContrastiveness + (1|LemmaParticiple) + (1|Text), pac, family = binomial)
```

The second model is the same as the first, but with the addition of varying slopes for lemma and text.

```r
costant.rs.both <- glmer(RealizationOfPAC ~ PreverbalContrastiveness + (PreverbalContrastiveness | LemmaParticiple) +
```

230
Figure 11.4: Individual by-text coefficients for glmer containing preverbal topicality information
CHAPTER 11. SEMANTIC AND PRAGMATIC PREDICTORS

|                | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 0.2109   | 0.1462     | 1.443   | 0.1490   |
| Topicalityyes  | -0.4362  | 0.2129     | -2.049  | 0.0404   |

Table 11.22: Final glmer for serialization ~ preverbal topicality

<table>
<thead>
<tr>
<th>Serialization</th>
<th>est factus (%)</th>
<th>factus est (%)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>noncontrastive</td>
<td>740 (36.19)</td>
<td>1305 (63.81)</td>
<td>2045</td>
</tr>
<tr>
<td>contrastive</td>
<td>219 (60.16)</td>
<td>145 (39.84)</td>
<td>364</td>
</tr>
<tr>
<td>Totals</td>
<td>959</td>
<td>1450</td>
<td>2409</td>
</tr>
</tbody>
</table>

Table 11.23: Relationship of preverbal contrastiveness and PAC serialization

(PreverbalContrastiveness|Text), pac, family=binomial

The third model contains varying slopes for only lemma:

contrast.rs.lemma<-glmer(RealizationOfPAC~PreverbalContrastiveness+
(PreverbalContrastiveness|LemmaParticiple)+
(1|Text), pac, family=binomial)

And the fourth model contains varying slopes for only text:

contrast.rs.text<-glmer(RealizationOfPAC~PreverbalContrastiveness+
(1|LemmaParticiple)+
(PreverbalContrastiveness|Text), pac, family=binomial)

When the models are compared, contrast.ri is the most parsimonious and should be preferred, as none of the other more complex models result in deviance reductions from 3062.99 that result in statistical significance on the relevant degrees of freedom. To exemplify, contrast.rs.lemma results in a deviance reduction of 2.6577 = 3062.99 − 3060.33, which on a χ²-distribution with 2 = 6 − 4 degrees of freedom is not significant (p = 0.2648).

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast.ri</td>
<td>4</td>
<td>3070.99</td>
<td>3094.14</td>
<td>-1531.49</td>
<td>3062.99</td>
</tr>
<tr>
<td>contrast.rs.lemma</td>
<td>6</td>
<td>3072.33</td>
<td>3107.05</td>
<td>-1530.17</td>
<td>3060.33</td>
</tr>
<tr>
<td>contrast.rs.text</td>
<td>6</td>
<td>3074.69</td>
<td>3109.42</td>
<td>-1531.35</td>
<td>3062.69</td>
</tr>
<tr>
<td>contrast.rs.both</td>
<td>8</td>
<td>3076.27</td>
<td>3122.57</td>
<td>-1530.14</td>
<td>3060.27</td>
</tr>
</tbody>
</table>

Let us therefore examine whether contrast.ri is a significantly more enhanced model in terms of the information the predictor adds compared to the baseline vcm (cf. Table 11.24). The inclusion of the predictor appears to be worth the additional complexity, as the model’s AIC is closer to zero (from 3139.41 to 3070.99) and the deviance has reduced significantly. Given its importance, we can now inspect the parameters for statistical interpretation (cf. Table 11.25). The level noncontrastiveness is the baseline level, and contrastive the level to be compared with it. The logit probability estimate is in terms of factus est (coded = 1). The coefficient indicates that when the preverbal constituent changes from being noncontrastive to being contrastive, the logit probability of factus est decreases by −1.0344, equating to an odds ratio of 0.3554396. Because the 95% confidence intervals do not include zero, lying entirely in negative logit space (2.5%: 0.7930317; 97.5%: 1.2757044), we can conclude that the population estimate is well reflected by the data sample. A further indication of its significance is the fact that the absolute z-value lies well above the critical z-value of 1.96 for large sample sizes. Contrastiveness plays a role in the alternation.

232
11.4. INFORMATION STRUCTURE

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline.vcm</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>contrast.ri</td>
<td>4</td>
<td>3070.99</td>
<td>3094.14</td>
<td>-1531.49</td>
<td>3062.99</td>
<td>70.42</td>
<td>1</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 11.24: ANOVA of baseline VCM and selected GLMER containing contrastiveness

| Estimate     | Std. Error | z value | Pr(>|z|) |
|--------------|------------|---------|----------|
| (Intercept)  | 0.3222     | 0.1613  | 1.997    | 0.0458   |
| PreverbalContrastivenesscontrastive | -1.0344 | 0.1231 | -8.400 | < 2e−16 |

Table 11.25: Final GLMER for serialization ~ contrastiveness

11.4.5 Discussion

Let us now turn to a global discussion of information structural effects. In sum, the effects are clearly in the direction as claimed by previous literature: the more focal/topic-worthy/new/contrastive the preverbal constituent, the more likely it is to stimulate an est factus order as opposed to a factus est one. The claims are therefore supported, and previous researchers were right to emphasize information structural effects. However, it is worth pointing out that such effects cannot be the “be-all-and-end-all” in describing the alternation. There are two facts that support this. The C-statistic which indicates predictive accuracy is well-below the requisite value of 0.80 for each of the selected models, particularly when compared to the respective baseline VCMs: definite.ri (0.66 vs. baseline 0.64), topicality.rs.text (0.65 vs. baseline 0.64), contrast.ri (0.73 vs. baseline 0.72). Second, partial pseudo-$R^2$ (Nagelkerke) measures for each of the information structural predictors, as assessed on basic single-level GLMs, reveal that none of them account for more than 5% of the variance in the response variable: definiteness accounts for 3.8%, topicality 0.4% and contrastiveness 4%. This suggests first that other predictors are clearly necessary to describe the variation and, most importantly, I believe that the literature has hitherto overemphasized the importance of such pragmatic factors. They are just a subset of the PAC’s information space, and their effects are rather weak, at least as operationalized here.

The robustness and reliability of the effects justifies, in my view, the use of proxy variables. Although definiteness, for example, is correlated with information status to a large extent, because the latter is not completely predictable from it (for instance, definite expressions, such as proper nouns, can be discoursally and/or situationally new), it is nonetheless clear that such surrogates are a very useful tool until such a time as we can tap information structure more accurately.

The final question that remains here is exactly why information structural concepts should impact on PAC choice. It seems fitting to mention two approaches: a syntactic feature-based approach and a complexity approach. Some syntactic models assume a discourse-configurational structure for language, where there are dedicated positions for information structural concepts (cf. e.g. Kiss 1998). In some approaches to generative grammar (e.g. the Minimalist Program of Chomsky, 1995), phrases can enter the syntax with unchecked features that need to be checked in the course of the derivation. Let us illustrate this with a simple example: Crassus Romam est profectus (non Fidenae) ‘Crassus departed for Rome (not Fidenae)’. Let us concentrate on the Romam est profectus string for syntactic analysis. The following small syntax tree in (71) comprises three main projections: the inflectional projection (IP), the verbal projection (VP), and a noun phrase (NP). The IP hosts the auxiliary as its head; the VP hosts the participle as its head; and the NP Romam is the directional complement of the verb. In this spirit of Minimalism, it can be assumed that the auxiliary, the t0, and the contrastive NP, Romam, are endowed with interpretable focus features, [iFoc].

23 Recall that the baseline C-values are different, because the baseline models for definite.ri and topicality.rs.text are based on only a subset of the data, specifically those with preverbal NP/PPs, whereas the contrastiveness model is based on the whole dataset and can be compared against the standard VCM.

24 Note that $R^2$ cannot be used for multilevel models.
To the IP a focus projection (FocP) is merged. This comprises a focus head (Foc\(^0\)) and a specifier (Spec,FocP), both of which have uninterpretable focus features, [uFoc], associated with them (cf. (72)).

The uninterpretable features need to be “checked”, otherwise the derivation will not converge. Uninterpretable features are checked by moving nodes with matching interpretable features into their domain. Thus, to check the uninterpretable feature on Foc\(^0\), the element in I\(^0\) docks into the former’s domain to check the feature. To check the uninterpretable feature on Spec,FocP, the NP with the interpretable focus feature, namely Romam moves into that position. Finally, the copies which are c-commanded by the moved entities go unpronounced at spell-out (cf. (73)).

The derivation is fine, as there are no uninterpretable features to check. However, such an analysis suffers from several drawbacks, one being whether we want to assume any type of configurational structure for Latin, and the second that such features would have to be specified with probabilities, which is completely ruled out in standard Chomskyan approaches to syntax.
Another approach, therefore, is to assume a flat structure to the Latin clause, and explain the linearization differences in terms of complexity. We have seen with previous variables that PACs with an unmarked or default morphology are typically serialized in a \textit{factus est} order. If one can extend the notion to context in general, the same notion of complexity may be at play here. If one assumes that VP-focus in the default (cf. Devine & Stephens 2006: 14), and narrow focus on a constituent other than the verb is a marked type of focus structure, then \textit{factus est} is used in the default context, and \textit{est factus} is used when more processing effort is required.

One could also explain the difference in order in terms of an optimization principle, as was seen for some of the phonological variables in Chapter 9. For example, to prevent two focal centres (one on the preverbal constituent and one on the lexical participle) from being adjacent, the informationally light auxiliary is brought into service as an informational buffer.

### 11.5 Summary remarks

This section has seen striking semantic and pragmatic effects bearing on the Latin PAC alternation, with the exception of animacy. To close, I summarize the effects in Table 11.26, which illustrates the predictor variables, the model selected, and the direction of the effects.

<table>
<thead>
<tr>
<th>Variable Group</th>
<th>Predictor</th>
<th>GLMER</th>
<th>Values/Levels favouring \textit{factus est}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Semantics</td>
<td>Eventuality Aspect</td>
<td>RI</td>
<td>achievement &gt; activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>accomplishment &gt; \textbf{non-eventive}</td>
</tr>
<tr>
<td></td>
<td>Verb Class</td>
<td>RI</td>
<td>appearance &gt; communication</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(motion) &gt; (possession)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(removing) &gt; (predicative complement)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(psych) &gt; (kill) &gt; \textbf{other}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(putting) &gt; (sending)</td>
</tr>
<tr>
<td>Animacy</td>
<td>Preverbal Animacy</td>
<td>VCM</td>
<td>\textbf{definite} &gt; indefinite</td>
</tr>
<tr>
<td>Information Structure</td>
<td>Definiteness</td>
<td>RI</td>
<td>\textbf{no} &gt; yes</td>
</tr>
<tr>
<td></td>
<td>Topicality</td>
<td>RS-TEXT</td>
<td>\textbf{noncontrastive} &gt; contrastive</td>
</tr>
</tbody>
</table>

Table 11.26: Summary of semantic and pragmatic predictors.
Chapter 12

Usage-based Predictors

12.1 Introduction

In this chapter we shall investigate the effects of two sets of usage-based variables, namely those concerning persistence and frequency. These two usage-based variables are two of the most common psycholinguistic variables mentioned in multivariate and probabilistic accounts of grammatical variation, so I include them here with respect to the Latin pac alternation. The chapter is structured more or less in the same way as preceding chapters, as are the sections. We begin with effect of persistence in Section 12.2 and then move on to frequency in Section 12.3. Section 12.4 offers a summary.

12.2 Persistence

12.2.1 Introduction

In this first section we discuss a well-established usage-based determinant of grammatical variation — persistence. When language users are first confronted with a choice between two or more semantically equivalent morphosyntactic variants in a discourse context, the variant they opt for on that particular occasion will often be exploited on a second occasion when they are next confronted with the same choice again. To exemplify with a concrete example, the writer in (1) is confronted with the genitive alternation in two slots, where they have to make a decision between the of-genitive \{the river systems of the continent, the Medical Association of the country\}, and the s-genitive \{the continent’s river systems, the country’s Medical Association\}. They opt for the s-genitive in both slots. The reason that the s-genitive […] occurs in the second slot instead of the other semantically equivalent of-genitive […] “might…be as simple as ‘because the speaker had just used that option…before’” (Szmrecsanyi, 2006: 209).

(1) In both countries the cases appeared to indicate what is most feared: that the continent’s river systems\[1\] are now infected, making the spread of the disease extremely difficult to control. In Ecuador, the country’s Medical Association\[2\] said 100 people had died of a total of 5,000 cases...

The phenomenon as illustrated in (1) goes by various terms, such as “copying”, “priming”, “parallelism”, “repetition”, “automaticity”, “persistence” (cf. Szmrecsanyi 2006; Hinrichs & Szmrecsanyi 2007: 455), and “perseveration” (e.g. Cameron & Flores-Ferrán 2004). In the following discussion, I shall adopt the term “persistence”, as used by, amongst others, Szmrecsanyi (2005, 2006), who defines it as:

---

1Example (1) is taken from Hinrichs & Szmrecsanyi (2007). In this particular example it is of note that a non-interchanging genitive \{the spread of the disease vs. *the disease’s spread\} intervenes, which seems to have no effect.
[T]he tendency that if speaker A faces a variable Z where he or she has the choice between two or more semantically equivalent variants..., speaker X’s choice will be affected by... previous exposure to the variable Z, such that use of a specific variant (either by speaker A or by another speaker B, to whose output speaker A has been exposed) in previous discourse will make it more likely, all other things being equal, that the same... variant will be used again by speaker A (2006: 2).

For simplicity, I will call the previous variable slot, e.g. the continent’s river systems[1] in (1), the “prime” and the upcoming variable slot, e.g. the country’s Medical Association[2] in (1) the “target”.

12.2.2 Evidence of persistence effects

12.2.2.1 General research

Robust effects of persistence have been reported in a now growing number of studies within the variationist arenas of sociolinguistics and corpus linguistics (e.g., Sankoff & Laberge 1978; Poplack 1980; Lefebvre 1981; Schiffrin 1981; Weiner & Labov 1983; Estival 1985; Mollica 1991; Scherre & Naro 1991, 1992; Poplack & Tagliamonte 1993; Cameron 1994; Labov 1994; Poplack & Tagliamonte 1996; Scherre 2001; Flores-Ferrán 2002; Cameron & Flores-Ferrán 2004; Travis 2005; Gries 2005; Szmercsanyi 2005, 2006; Travis 2007; Bresnan et al. 2007; Hinrichs & Szmercsanyi 2007; Bresnan & Ford 2010; Gries & Hilpert 2010; de Marneffe et al. 2012; amongst others).

To give several examples, Weiner & Labov (1983) explore the language-external and language-internal constraints on the alternation between the agentless passive, e.g. a watch got/was stolen, vs. the generalized active, e.g. someone stole a watch, basing themselves on 1489 datapoints. Two important points should be reported here. First, they find that if a passive precedes the target variable site within the last five clauses, then the use of an agentless passive in the target is encouraged: specifically, 72% of datapoints that are preceded by a passive are realized as a passive, whereas the baseline average for an agentless passive is 35%. Second, they find that if the logical object of the clause that contains the target site is co-referential with another NP in the preceding span of discourse that is realized in a subject position, then a passive structure is effected — thus copying the position of the subject to the target clause. When the coreferential NP is realized in another position, the use of the passive in the target is discouraged. Concluding, they note that the “mechanical tendency to preserve parallel structure” is a strong influence (Weiner & Labov, 1983: 56).

Szmercsanyi (2006)2, the most rigorous study to date, discusses persistence effects in the previously-studied particle alternation (V NP Ptc1 vs. V Ptc1 NP), and four new alternations: comparison choice (analytic vs. synthetic, e.g. more happy vs. happier), genitive choice (s-genitive vs. of-genitive), future marker choice (be going to vs. will), and complementation choice (gerund vs. to-infinitive, e.g. begin doing vs. begin to do). He uses as his corpus the British National Corpus (BNC) and Freiburg English Dialect Corpus (FRED), with datasets encompassing a vast number of datapoints (e.g over 100,000 with respect to the future marker alternation). Persistence effects are found to be significant parameters in the logistic regression models for all alternations. More importantly, inclusion of such variables robustly increase the fit and accuracy of the models.3 Besides persistence main effects, persistence participates in a number of significant interactions — for instance distance from the current variable site to the previous variable site (measured logarithmically) is shown to be “perhaps the most important” moderator of persistence (2006: 185): “persistence declines significantly after a relatively short period after a choice has been made (10 words, or 5 seconds of talk), but on the whole it is fairly long-lived before the effect dissipates entirely” (2006: 190). Finally, Szmercsanyi also finds that another important moderator of persistence is a matching verb lemma, much like Gries (2005): if the previous variable site is similar in

---

2This book expands in detail on Szmercsanyi (2005).
3See above for information on variables that may be significant but not add explanatory power to the model.
lexis as the upcoming variable site, then the latter is more likely to be of the same structure too: thus, e.g., a previous he turned the light off is more likely to stimulate he turned his laptop off than he picked his laptop up because of the match in verbal lexical material.

12.2.2.2 AVc research

We turn now to persistence in the AVc literature specifically. Hartsuiker & Westenberg (2000) is a psycholinguistic investigation into the effect of persistence in the domain of Dutch verb clusters. In Dutch, verb clusters containing a participle and an auxiliary alternate — in the case of a subordinate clause environment — between a participle-first order, as in (2), and a participle-final order, as in (3).

(2) De man belde de politie omdat zijn portemonnee gestolen was
(Hartsuiker & Westenberg (2000: B31, ex. 10))

(3) De man belde de politie omdat zijn portemonnee was gestolen
(Hartsuiker & Westenberg (2000: B31, ex. 9))

Specifically, Hartsuiker & Westenberg (2000) were interested in whether the word order of the participle and the auxiliary in a “target” subordinate clause was influenced by the word order of these elements in a previously used “prime” subordinate clause. For instance, if the participle-first serialization was used in the priming context, as in say (2), to what extent does this influence the same participle-first order gebroken had in the upcoming target context (4) as opposed to the participle-last variant had gebroken?

(4) De skier lag in het ziekenhuis omdat hij zijn been {had gebroken, gebroken had}
(cf. Hartsuiker & Westenberg (2000: B34, Table 1))

The experiment, which makes use of 66 native speakers of Dutch as subjects, is split so that effects in both written and spoken conditions can investigated. The subjects are asked to first complete a prime context, which is “constrained, so as to yield a sentence with the intended syntactic structure” (2000: B31). Hence, in (5), the context allows only an auxiliary, resulting in a participle-auxiliary prime; in (6) it allows only a participle, resulting in an auxiliary-participle prime.

(5) Ik kon er niet door omdat de weg geblokkeerd {was}
(cf. Hartsuiker & Westenberg (2000: B31))

(6) Ik kon er niet door omdat de weg was {geblokkeerd}
(cf. Hartsuiker & Westenberg (2000: B31))

Subjects are then asked to complete the target context, which is “less constrained” in that a verb cluster should be realized, but its specific realization is essentially free, as in (7), which can allow the participle-first serialization gezien had, or the participle-final serialization had gezien.

(7) Jan vertelde de inspecteur dat hij niets {had gezien, gezien had}

They find significant persistent effects in both modalities, concluding convincingly (Hartsuiker & Westenberg, 2000: B36) that:

a determinant of the word order of participle and auxiliary verbs in Dutch subordinate clauses is the word order of a sentence immediately produced before. Following prime fragments that elicited an auxiliary-final word order, there were more completions with that word order than following a prime which elicited the alternative, participle-final, word order.

Drawing explicitly on these results obtained by Hartsuiker & Westenberg (2000), De Sutter (2005, 2009) investigates the effect of persistence on the serialization of the participle and the auxiliary in
journalistic texts written in Belgian Dutch (\(N = 2390\)). He operationalizes persistence by scanning the previous ten clauses before the verb cluster to determine if there is a verb cluster or not, and if there is, whether it is participle-first, e.g. *gesloten is* ‘closed is’ or participle-final, e.g. *is gesloten* ‘is closed’ (2009: 243). When this variable is pitted against other potential predictors in a generalized linear model (multiple logistic regression), he finds that persistence exerts a highly significant effect on the internal serialization of the verb cluster. More specifically, his results show when the last verb cluster is realized as participle-final, the odds of a participle-final order being realized in the target clause are 3.28 times greater than when the last verb cluster is realized participle-initially (2009: 245). Thus, these results fit squarely with those of Hartsuiker & Westenberg (2000) and, I note, more generally with the persistence literature, discussed above. In addition, De Sutter finds that when there is no cluster in the preceding ten-clause context, i.e. when the speaker (probably) has to make a fresh decision as to which serialization option to go for, the odds for the participle-first order are 1.72 times higher than when the last verb cluster is participle-first (De Sutter, 2009: 245). This is perhaps unsurprising, given that the participle-final order is the preferred variant in the dataset taken as a whole (66.99%) (2009: 235).

To the best of my knowledge, nothing has been said of the phenomenon for Latin *pacs*. This work aims to remedy this.

12.2.3 Variable profile and operationalization

The operationalization of this variable is straightforward, coded as the realization of the previous response token, i.e. *fe* for *factus est* and *ef* for *est factus*. Where there is no previous realization, i.e. when the datapoint is at the beginning of the text, a missing datapoint is coded (NA). In sum, there are 9 missing datapoints for this variable, i.e. one for each of the texts (cf. Table 12.1 for the first six observations).

<table>
<thead>
<tr>
<th>Observation</th>
<th>Target Realization</th>
<th>Prime Realization</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>parati.essent</td>
<td>fe</td>
</tr>
<tr>
<td>2</td>
<td>est.consecutus</td>
<td>ef</td>
</tr>
<tr>
<td>3</td>
<td>erat.mandatum</td>
<td>ef</td>
</tr>
<tr>
<td>4</td>
<td>erant.signatae</td>
<td>ef</td>
</tr>
<tr>
<td>5</td>
<td>est.uiura</td>
<td>ef</td>
</tr>
<tr>
<td>6</td>
<td>esset.occupatus</td>
<td>ef</td>
</tr>
</tbody>
</table>

Table 12.1: Example coding of the persistence predictor

12.2.4 Statistical modelling

The overall distribution of prime and target is reported in Table 12.2. There appears to be an effect in the expected direction, although it is noticeably weak (\(\chi^2 = 13.97, df = 1, p = 0.0001859\)), with *factus est* primes stimulating proportionally more *factus est* serializations in the target slot and *est factus* primes stimulating proportionally more *est factus* serializations in the target slot.

<table>
<thead>
<tr>
<th></th>
<th>Target</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>est factus</em> (%)</td>
<td><em>factus est</em> (%)</td>
<td></td>
</tr>
<tr>
<td><em>est factus</em> prime</td>
<td>426 (44.47)</td>
<td>532 (55.53)</td>
<td></td>
</tr>
<tr>
<td><em>factus est</em> prime</td>
<td>530 (36.75)</td>
<td>912 (63.25)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>956</td>
<td>1444</td>
<td></td>
</tr>
</tbody>
</table>

Table 12.2: Distribution of prime and target

However, before we can statistically model this distribution, a variance components model specific to this variable is first constructed, given that we are 9 observations short of the full dataset and should not therefore use the baseline variance components model established in Chapter 8 for comparison. The first
model to be constructed is the basic single-level GLM. Table 12.3 shows the deviance and AIC values for several models considered. All three VCMS are significantly better than a null GLM which only has access to the base intercept. Further, the two individual nested VCMS are significantly worse than the more complex VCM containing both lemmatic and textual information. Specifically, the model containing just lemma is significantly poorer than the model that contains both lemmatic and textual information ($\chi^2 = 94.162$, df = 1, $p < 2.2e^{-16}$), and the model containing just text is also significantly poorer than the model that contains both lemmatic and textual information ($\chi^2 = 11.768$, df = 1, $p = 0.0006025$). As such, we accept the last model in Table 12.3 as the baseline model with which to compare persistence effects.

We now begin to construct a model for statistical interpretation. Several models are constructed and compared. First, a random intercepts model is constructed with the persistence predictor; the R code is:

```r
persistence.ri <- glmer(RealizationOfPAC ~ (1 | LemmaParticiple) + (1 | Text) + PreviousRealization, pac, family = binomial)
```

Second, a random slopes model is constructed for lemma and text; the R code is:

```r
persistence.rs.both <- glmer(RealizationOfPAC ~ (PreviousRealization | LemmaParticiple) + (PreviousRealization | Text) + PreviousRealization, pac, family = binomial)
```

Third, a random slopes model is constructed for just lemma; the R code is:

```r
persistence.rs.lemma <- glmer(RealizationOfPAC ~ (PreviousRealization | LemmaParticiple) + (1 | Text) + PreviousRealization, pac, family = binomial)
```

Fourth, a random slopes model is constructed for just text. The R code for this model is:

```r
persistence.rs.text <- glmer(RealizationOfPAC ~ (1 | LemmaParticiple) + (PreviousRealization | Text) + PreviousRealization, pac, family = binomial)
```

These models are contrasted and compared by means of analysis of deviance statistics. The salient ANOVA output for these four models is presented in Table 12.4. As we can see, all the models are non-significant from the model that includes just the persistence predictor and random intercepts of verb lemma and text. That is, there are no significant interactions between persistence and the verb lemma or between persistence and the text source. In other words, persistence appears to be operating very similarly, regardless of the verb lemma involved or the text. Consequently, the model termed `persistence.ri` is taken up for further appraisal because of its lowest AIC (all the models have roughly the same amount of deviance). Finally, comparing this latter model to the baseline VCM above, we see that the selected model is actually inferior. The AIC of the VCM is 3126.708, while that of `persistence.ri` is 3126.80, which are both very similar. In addition, the reduction of deviance of $1.908 = 3120.708 - 3118.80$ on $1 = 4 - 3$ degrees of freedom does not reach significance. In sum, it appears that persistence plays no real effect in the present dataset, despite the significant result of the $\chi^2$-test above. This is confirmed by checking the parameters for `persistence.ri` in Table 12.5. According to this output, the estimated coefficient of 0.12743 is non-significant when it divided by its standard error ($p = 0.159$). Thus, the more parsimonious model should be selected, which includes only the varying intercepts of text and verb lemma.

### Table 12.3: Constructing a baseline VCM for persistence effects

<table>
<thead>
<tr>
<th>Model</th>
<th>$-2LL$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM null</td>
<td>3227.184</td>
<td>3229.184</td>
</tr>
<tr>
<td>VCM + varying lemma intercepts</td>
<td>3214.870</td>
<td>3218.870</td>
</tr>
<tr>
<td>VCM + varying text intercepts</td>
<td>3132.476</td>
<td>3136.476</td>
</tr>
<tr>
<td>VCM + varying lemma and text intercepts</td>
<td>3120.708</td>
<td>3126.708</td>
</tr>
</tbody>
</table>

240
12.2. PERSISTENCE

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>persistence.ri</td>
<td>4</td>
<td>3126.80</td>
<td>3149.94</td>
<td>-1559.40</td>
<td>3118.80</td>
</tr>
<tr>
<td>persistence.rs.lemma</td>
<td>6</td>
<td>3130.79</td>
<td>3165.49</td>
<td>-1559.40</td>
<td>3118.79</td>
</tr>
<tr>
<td>persistence.rs.text</td>
<td>6</td>
<td>3130.63</td>
<td>3165.33</td>
<td>-1559.32</td>
<td>3118.63</td>
</tr>
<tr>
<td>persistence.rs.both</td>
<td>8</td>
<td>3134.61</td>
<td>3180.87</td>
<td>-1559.30</td>
<td>3118.61</td>
</tr>
</tbody>
</table>

Table 12.4: ANOVAs for persistence models

|                        | Estimate | Std. Error | z value | Pr(>|z|) |
|------------------------|----------|------------|---------|---------|
| (Intercept)            | 0.10143  | 0.16204    | 0.626   | 0.531   |
| PreviousRealizationfe | 0.12743  | 0.09039    | 1.410   | 0.159   |

Table 12.5: Parameters for persistence model

However, it would be unwise to immediately dismiss persistence as a predictor. First, it has been argued that persistence is more effective when the target’s lemma is identical to the priming lemma (cf. Szmrecsanyi 2006). A further possibility is that persistence might be more effective when the target auxiliary is identical to the priming auxiliary. We will now explore these possibilities in greater detail.\(^4\)

Let us begin with identical lemma effects. An \( R \) script was configured to establish for each observation whether the target observation and the prime had the same or a different participle lemma. In (8) the lemmas across the prime and the target are identical (they are both from the lemma \( \text{pugno} \) ‘fight’), while in (9) they are different (the prime is from the lemma \( \text{refero} \) ‘report’, whilst the target is from the lemma \( \text{proficiscor} \) ‘depart’).

(8) \( \text{diu atque acriter pugnatum est... cum ab hora long and fierce fight} \).

(9) \( \text{uerum falsum=ne sibi esset report.} \).

Again, four models are constructed with respect to this interaction effect. When interaction effects are entered into a model, it is customary to include both main effects and the interaction rather than just including the interaction term. We are therefore trying to establish whether the main effects are important as well as their interaction. The first model includes predictors for persistence predictor and the prime’s lexical similarity and their interaction, as well as random intercepts of lemma and text. Its \( R \) code is as follows:

```r
per.verb.ri<-glmer(Serialization~PreviousSerial*PreLemmaSim+(1|Lemma)+(1|Text), data=per,family=binomial)
```

We now introduce varying slopes. In order to keep the model building process straightforward, varying slopes for only the persistence predictor are included. The second model therefore includes predictors for

\(^4\)A third possibility is that persistence effects are more striking when the prime and the target are in closer proximity, as has been shown by e.g. Szmrecsanyi (2006). However, this is difficult to code using the current text archive, so I leave this possibility for further investigation.
persistence predictor and the prime’s lexical similarity and their interaction, as well as random intercepts of lemma and text and random slopes for lemma.

\[ \text{per.verb.rs.lemma}<-\text{glmer(Serialization}^\text{PreviousSerial*PreLemmaSim}+(\text{PreviousSerial}\mid\text{Lemma})+(1\mid\text{Text}), \text{data=}\text{per, family=}\text{binomial}) \]

The third model includes predictors for persistence predictor and the prime’s lexical similarity and their interaction, as well as random intercepts of lemma and text and random slopes for text.

\[ \text{per.verb.rs.text}<-\text{glmer(Serialization}^\text{PreviousSerial*PreLemmaSim}+(1\mid\text{Lemma})+(\text{PreviousSerial}\mid\text{Text}), \text{data=}\text{per, family=}\text{binomial}) \]

Finally, the fourth model includes predictors for persistence predictor and the prime’s lexical similarity and their interaction, as well as random intercepts of lemma and text and random slopes for both lemma and text.

\[ \text{per.verb.rs.both}<-\text{glmer(Serialization}^\text{PreviousSerial*PreLemmaSim}+(\text{PreviousSerial}\mid\text{Lemma})+(\text{PreviousSerial}\mid\text{Text}), \text{data=}\text{per, family=}\text{binomial}) \]

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>per.verb.ri</td>
<td>6</td>
<td>3128.96</td>
<td>3163.66</td>
<td>-1558.48</td>
</tr>
<tr>
<td>per.verb.rs.lemma</td>
<td>8</td>
<td>3132.95</td>
<td>3179.21</td>
<td>-1558.47</td>
</tr>
<tr>
<td>per.verb.rs.text</td>
<td>8</td>
<td>3132.77</td>
<td>3179.03</td>
<td>-1558.38</td>
</tr>
<tr>
<td>per.verb.rs.both</td>
<td>10</td>
<td>3136.74</td>
<td>3194.57</td>
<td>-1558.37</td>
</tr>
</tbody>
</table>

Table 12.6: Model comparisons for persistence and prime lemma models

These models are then contrasted and compared by means of analysis of deviance statistics. The salient ANOVA output for these four models is presented in Table 12.6. Going on the AIC information, it appears that the least complex model, per.verb.ri, is the most parsimonious. However, when this model is compared against the baseline vcm model (the fourth model listed in Table 12.3), the AIC value is higher in the former: the AIC of per.verb.ri is 3126.80, while that of the baseline vcm is 3126.7. The difference in deviance \(3.748 = 3120.708 - 3116.96\) on \(1 = 4 - 3\) degrees of freedom is only marginally significant \((p = 0.05287074)\). As such, we revert to the baseline vcm model as being the most parsimonious. In conclusion, then, persistence effects are not stronger when the lemma of the prime and the target is identical for the PAC alternation.

Let us now turn to the interaction between persistence and the similarity of the prime’s auxiliary and the target’s auxiliary. As with the preceding variable, an R script was configured to establish for each observation whether the target observation and the prime had the same or a different auxiliary word form. For instance, in (10) the auxiliaries across the prime and the target are identical (they are both est), while in (11) they are different (the prime is sunt, whilst the target is est).

(10) frumentatum est profectus. ex eo
    forage.SPN be.BE.3.SG.PRS.IND depart.PTCP.PL.REF.M.SG.NOM. from PRN.N.SG.ABL
    est cognitum Caesar. consilium
    be.BE.3.SG.PRS.IND find.out.PTCP.PL.REF.N.SG.NOM Caesar.M.SG.GEN plan.N.SG.NOM
    illum PRN.N.SG.NOM
    ‘...departed to forage. As a result of that Caesar’s plan was discovered...’ (B. Afr. 11. 3)
Again, four models are constructed with respect to this interaction effect. The first model includes predictors for persistence predictor and the prime’s auxiliary similarity and their interaction, as well as random intercepts of lemma and text. Its R code is as follows:

```r
per.aux.ri<-glmer(Serialization~PreviousSerial*PreAuxSim+(1|Lemma)+(1|Text),
data=per,family=binomial)
```

We now introduce varying slopes. As before, in order to keep the model building process straightforward, varying slopes for only the persistence predictor are included. The second model therefore includes predictors for persistence predictor and the prime’s auxiliary similarity and their interaction, as well as random intercepts of lemma and text and random slopes for lemma.

```r
per.aux.rs.lemma<-glmer(Serialization~PreviousSerial*PreAuxSim+(PreviousSerial|Lemma)+(1|Text),
data=per,family=binomial)
```

The third model includes predictors for persistence predictor and the prime’s auxiliary similarity and their interaction, as well as random intercepts of lemma and text and random slopes for text.

```r
per.aux.rs.text<-glmer(Serialization~PreviousSerial*PreAuxSim+(1|Lemma)+(PreviousSerial|Text),
data=per,family=binomial)
```

Finally, the fourth model includes predictors for persistence predictor and the prime’s auxiliary similarity and their interaction, as well as random intercepts of lemma and text and random slopes for both lemma and text.

```r
per.aux.rs.both<-glmer(Serialization~PreviousSerial*PreAuxSim+(PreviousSerial|Lemma)+(PreviousSerial|Text),data=per,family=binomial)
```

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>per.aux.ri</td>
<td>6</td>
<td>3105.59</td>
<td>3140.29</td>
<td>-1546.79</td>
</tr>
<tr>
<td>per.aux.rs.lemma</td>
<td>8</td>
<td>3109.59</td>
<td>3155.85</td>
<td>-1546.79</td>
</tr>
<tr>
<td>per.aux.rs.text</td>
<td>8</td>
<td>3109.37</td>
<td>3155.64</td>
<td>-1546.69</td>
</tr>
<tr>
<td>per.aux.rs.both</td>
<td>10</td>
<td>3113.37</td>
<td>3171.20</td>
<td>-1546.69</td>
</tr>
</tbody>
</table>

Table 12.7: Model comparisons for persistence and prime’s auxiliary similarity interaction

Inspecting the comparisons of these models in terms of their deviance and AIC in Table 12.7, we see that all four models have roughly the same deviance. However, in terms of AIC values, the least complex model wins the day, namely `per.aux.ri`, which includes both the main effects, the interaction between the two main predictors, and random intercepts for text and lemma. Specifically, its AIC is 3105.59 which is the closest to zero out of the four models up for comparison. Let us now compare this model with the baseline `vcm`, information which is presented in Table 12.8. As this table shows, the AIC is closer to zero in the more informative model (3105.59 as compared with 3126.71) and furthermore the reduction in deviance is highly significant on 3 degrees of freedom (3093.59 as compared with 3120.71) Consequently, it appears that there is some effect.
CHAPTER 12. USAGE-BASED PREDICTORS

|                | Df | AIC   | BIC   | logLik | deviance | Chiq  | Chi Df | Pr(|Chiq|) |
|----------------|----|-------|-------|--------|----------|-------|--------|---------|
| per.glmer.both | 3  | 3126.71 | 3144.06 | -1560.35 | 3120.71 |       |        |         |
| per.aux.ri    | 6  | 3105.59 | 3140.29 | -1546.79 | 3093.59 | 27.12 | 3      | 0.0000  |

Table 12.8: ANOVA of baseline VCM and GLMER containing persistence and prime’s auxiliary similarity information

The nature of the effect is reported in Table 12.9. It establishes that while there are no significant main effects of either persistence or of the prime’s auxiliary similarity, there is indeed a significant interaction between the two. Specifically, the fourth line of the output indicates that when the prime is *factus est* and the prime’s auxiliary is the same as that in the target slot, then the odds of a *factus est* outcome increase by \( \exp(0.66469) = 1.94 \) times when it is not. This is clearly in the direction that we would expect if persistence had an effect. When the estimate is divided by its standard error, we obtain a z-score of 3, which is greater than the critical z-score of 1.96 for this large sample size.

|                                | Estimate | Std. Error | z value | Pr(|z|) |
|--------------------------------|----------|------------|---------|--------|
| (Intercept)                    | 0.10822  | 0.16344    | 0.662   | 0.50789|
| PreviousSerialfe               | -0.04092 | 0.10166    | -0.402  | 0.68731|
| PreAuxSimsame                  | 0.01819  | 0.17287    | 0.105   | 0.91618|
| PreviousSerialfe:PreAuxSimsame | 0.66469  | 0.22087    | 3.009   | 0.00262|

Table 12.9: GLMER of persistence and auxiliary similarity

12.2.5 Discussion

In this section we have explored persistence as a predictor in the information space of the PAC alternation. While no independent effects of persistence were found, an interaction effect was discovered between the persistence predictor and the similarity of the auxiliary in the prime and the target. When the prime is *factus est* and the auxiliary used in the prime and the target is identical, then the effect of persistence is enhanced, thus making it more likely to serialize the PAC as *factus est*.

It is worth pondering why no main effect of persistence was found, particularly given its prominence in the literature noted in Section 12.2.2. There are, I believe, several reasons for this. First, the modality here is written rather than spoken. It may be that persistence has a more robust effect in spoken language than in the written form. Since written text is often edited and can be produced over longer periods of time, the persistence effect is more likely to be diluted. That said, this does not seem to have prevented main effects being observed in the Dutch dataset of De Sutter (2009), which was based on journalistic texts. A second possibility is that structural persistence has more of an effect in languages that display hierarchical structure as opposed to languages that have flat, or non-configurational, structure. It has been argued that Latin does not have much in the way of hierarchical structure, except for in the CP and PP domains (cf. Ledgeway 2012). These are mere speculations as to why no main effect of persistence was located. It will be up to further research to scrutinize these speculations in detail.

12.3 Frequency

12.3.1 Introduction

In this section we discuss the second usage-based variable group — the potential effect of frequency related predictors.
12.3. FREQUENCY

12.3.2 Background

12.3.2.1 General research

In general terms the presence of frequency related effects is conceptually well motivated. There is a wealth of evidence that frequency of use and exposure impinges on language and language structures (Bybee & Hopper 2001).

A lexical item's frequency of occurrence has a facilitative effect on its accessibility and comprehension (Jurafsky, 2003: 44). In psycholinguistic studies, fixation durations are shorter if the lexical item is more frequent (Jurafsky, 2003: 44; cf. Inhoff, 1984). In word recognition tasks, the more frequent the lexical stimulus, the shorter the recognition time (Jurafsky, 2003: 44; cf. Howes & Solomon, 1951; Rubenstein et al., 1971). In naming tasks, higher frequency words are read out more quickly (Jurafsky, 2003: 22; cf. Forster & Chambers, 1973). In neurolinguistic research, the more frequent the content word of a lexical stimulus, the lower its negative amplitude in the N400 component (see Bornkessel-Schlesewsky & Schlesewsky, 2009: 8 for discussion and references).

In phonology, frequency is of central importance. As first observed by Hugo Schuchardt (Schuchardt, 1885), sound change operates on higher frequency words first (Hooper, 1976; Jaeger, 2010). The literature has focused on various types of phonological reduction and deletion as influenced by frequency. Bybee reports on post-stress schwa deletion in [VCVry] sequences such as every and memory (Hooper, 1976). She finds that deletion is heavily influenced by the frequency of the lexical item: family is a high frequency word and has its schwa deleted; artillery, by contrast, is a low frequency word and the schwa is retained. Similar effects have been reported for t/d deletion in words such as and and meant (Bybee, 2000). The former, being highly frequent, regularly has its final dental dropped, whereas meant, a word of low frequency, retains it.

Frequency is relevant to morphosyntactic variation, but not in the same way as phonological variation. In this domain, high frequency words and sequences resist innovative morphosyntactic patterns and levelling (Hooper, 1976). Bybee & Thompson (1997) point out some further interesting cases of frequency bearing on grammar. Pronouns and noun phrases behave differently with respect to case marking, for instance, because the former are more frequent. Pronouns also show peculiarities of word order with respect to the verb. In Spanish, object pronouns are proclitic to the verb, whereas full noun phrases are postverbal. The former's OV order is argued to be the older pattern, which has become engrained structurally through repeated use. In Canadian French, embedded complement clauses can be variably realized with indicative or subjunctive. Realizations in the subjunctive occur in highly frequent contexts, where the matrix verb is falloir and the subordinated verb is a verb of a high frequency, such as aller, avoir, and être.

12.3.2.2 AVC research

In addition to the above studies that establish reliable effects of frequency, there are construction-specific motivations for including frequency. For instance, frequency has also been discussed in quantitative studies of alternation within AVCs. De Sutter (2009), writing on Belgian Dutch AVCs (e.g. gesloten is 'closed is' vs. is gesloten 'is closed'), embraces a remark by Barlow & Kemmer (2000: x), who state that “[h]igher frequency of a unit or pattern results in a greater degree of what Langacker terms entrenchment, i.e. cognitive routinization, which affects the processing of a unit”. Accordingly, he includes the frequency of the participle in his variable portfolio. When it comes to the logistic regression, he finds that when the frequency of the participle is higher, the likelihood of the participle-final word order is increased. It is not stated how this should be interpreted, but there is a brief inconclusive discussion in his 2005 thesis.

Be that as it may, there has been absolutely no discussion of the effect of frequency concerning the PAC alternation, and this is the first study to examine its effect in this arena.
12.3.3 Variable profile and operationalization

Against this backdrop, I investigate how frequency might influence serialization choice in PACS. Two frequency predictors are considered in this study, namely that of the participle’s word form (FreqPart) and that of the auxiliary’s word form (FreqAux). Frequency information for these two variables was obtained from LLT-A, the text archive on which the present dataset is based; the Antiquitas portion of the text archive was used.

These are two metric predictors (or at least they are to begin with). For the frequency of the participle, the range of frequencies is from 1 to 2179 (mean = 236.988, median = 79, standard deviation = 413.0935). For the frequency of the auxiliary, the range of frequencies is from 25 to 67024 (mean = 31727.55, median = 17336, standard deviation = 27532.24). The univariate distribution for each variable is given in Figure 12.1. As the differences between the means and medians suggest and as these kernel density plots show, both variables are evidently not normally distributed: the frequency of the participle exhibits a positive skew, as it is skewed to the right, whereas the frequency of the auxiliary appears to be bimodal with two peaks.

![](image1.png)

Figure 12.1: Kernal density plots for frequency of participle (top) and frequency of auxiliary (bottom)

12.3.4 Statistical modelling

12.3.4.1 Frequency of the participle

We begin with the frequency of the participle. Because this is a metric predictor, we first need to assess whether the linearity in the logit assumption is met. To establish the precise relationship between the logit probability of factus est and the frequency of the participle, a series of generalized additive models
(GAMS) were fitted to the data. GAMS are excellent at showing the underlying relationship between the two variables (Crawley, 2012). Four models are fitted. In the first model the raw frequency information is used. In the second model frequency is logarithmically transformed (natural log). In the third model a quadratic polynomial term is fitted. In the fourth a cubic polynomial term is fitted. The non-parametric smoothing lines are plotted in Figure 12.2. Raw frequency is roughly linear with respect to the logit probability, but it is the log transformed frequency that is the most strikingly linear. Another possibility is to bin the observations in terms of low frequencies and high frequencies. Using the ctree function in the party package, a significant split is established at a count of \( \leq 123 \). With three basic GLMs fitted to the data, the model containing the binned frequency predictor has the lowest deviance (3214.478 vs. 3220.2 (log) vs. 3222.2 (raw)) and AIC closer to zero (3218.478 vs. 3224.2 (log) vs. 3226.2 (raw)). The binning is also conceptually motivated, as well as statistically, because it has the benefit of separating the participles in the peak of the kernel density plot from those that straddle in the long right tail. Consequently, we will work with the binary binned predictor in this instance.

![Figure 12.2: Non-parametric smoothers for frequency of participle](image)

With the nature of the relationship established, we will now start to build the appropriate model for interpretation. As usual, four models are constructed and compared. The first model contains the binary binned frequency predictor along with varying intercepts for verb lemma and text.

\[
\text{freqpart.ri<-glmer(RealizationOfPAC~(1|LemmaParticiple) + (1|Text)+BinnedFreqPart,pac,family=binomial)}
\]

The next three models contain varying slopes in addition to varying intercepts. The second model therefore contains the binned frequency predictor along with varying intercepts for verb lemma and text, as well as varying slopes for the verb’s lemma.
freqpart.rs.lemma<-glmer(RealizationOfPAC~(BinnedFreqPart|LemmaParticiple) +(1|Text)+BinnedFreqPart,pac,family=binomial)

The third model contains the predictor along with varying intercepts for verb lemma and text, as well as varying slopes for text.

freqpart.rs.text<-glmer(RealizationOfPAC~(1|LemmaParticiple) +(BinnedFreqPart|Text)+BinnedFreqPart,pac,family=binomial)

Finally, the fourth model contains the predictor with varying intercepts and slopes for both verb lemma and text.

freqpart.rs.both<-glmer(RealizationOfPAC~(BinnedFreqPart|LemmaParticiple) +(BinnedFreqPart|Text)+BinnedFreqPart,pac,family=binomial)

The deviance ($-2LL$) and $AIC$ values (along with related statistics) are reported in Table 12.10. When each of the three more complex models (freqpart.rs.lemma, freqpart.rs.text, freqpart.rs.both) are compared with the more simple varying intercepts model (freqpart.ri), freqpart.rs.lemma ($\chi^2 = 9.9688$, $df = 2$, $p = 0.006844$) and freqpart.rs.both ($\chi^2 = 9.9915$, $df = 4$, $p = 0.04057$) are significantly better models, whereas freqpart.rs.text is not ($\chi^2 = 0.0629$, $df = 2$, $p = 0.969$). A further ANOVA comparison of freqpart.rs.lemma and freqpart.rs.both confirms that only the former is warranted, as the inclusion of slopes for both the multilevel terms does not significantly reduce the deviance ($\chi^2 = 0.0227$, $df = 2$, $p = 0.9887$). Note further that the former model has the lowest $AIC$ of 3114.54. Therefore, the freqpart.rs.lemma model is adopted for further evaluation.

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>freqpart.ri</td>
<td>4</td>
<td>3120.51</td>
<td>3143.66</td>
<td>-1556.25</td>
</tr>
<tr>
<td>freqpart.rs.lemma</td>
<td>6</td>
<td>3114.54</td>
<td>3149.26</td>
<td>-1551.27</td>
</tr>
<tr>
<td>freqpart.rs.text</td>
<td>6</td>
<td>3124.45</td>
<td>3159.17</td>
<td>-1556.22</td>
</tr>
<tr>
<td>freqpart.rs.both</td>
<td>8</td>
<td>3118.52</td>
<td>3164.81</td>
<td>-1551.26</td>
</tr>
</tbody>
</table>

Table 12.10: Deviance and related statistics

The next step is to compare this model with the baseline vcm. An ANOVA comparing the two is given in Table 12.11. As we can see, the reduction in deviance of 30.87 = 3133.41 - 3102.54 is highly significant on 3 = 6 - 3 degrees of freedom ($p = 9.066625e-07$). There are thus important effects to be found with respect to the frequency of the participle.

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>glmer.vcm.both</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>freqpart.rs.lemma</td>
<td>6</td>
<td>3114.54</td>
<td>3149.26</td>
<td>-1551.27</td>
<td>3102.54</td>
<td>30.87</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 12.11: ANOVA of baseline vcm and glmer containing participle frequency information

Let us now turn to an interpretation of the parameters of the adequate model, reported in Table 12.12. When the frequency of the participle increases by one unit, that is when it changes from a low frequency participle (dummy coded 0) to a high frequency participle (dummy coded 1), the logit probability of a factus est outcome increases by 0.39477. Translated into odds, this means that the likelihood of a factus est outcome increases by a factor of 1.484043. The 95% confidence intervals associated with the logit estimate range from 0.205724 to 0.5837573 (or in terms of odds ratios from 1.228485 to 1.792762). Given that the confidence intervals do not include zero in the logit or 1 in the odds, and given that the estimate divided by the standard error 0.39477/0.09642 = 4.094 gives a z-statistic greater than the large sample critical value of 1.96, we can conclude that it is a highly significant parameter estimate. In gross terms, then, the more frequent the participle, the more likely it is to be realized as factus est. Why this might be so is discussed in Section 12.3.5. Before we do so, we turn to an analysis of the auxiliary’s frequency.
12.3. FREQUENCY

| Estimate  | Std. Error | z value | Pr(>|z|) |
|-----------|------------|---------|----------|
| (Intercept) | 0.06402 | 0.16237 | 0.394 | 0.693 |
| BinnedFreqParthigh | 0.39477 | 0.09642 | 4.094 | 4.24e-05 |

Table 12.12: glmer of serialization ~ participle frequency

12.3.4.2 Frequency of the auxiliary

As we did for the frequency of the participle, we split the auxiliary frequency into two bins, low frequency and high frequency. A significant split is found between auxiliary frequencies $\leq 25227$ and those of a higher frequency. This binning is again conceptually better motivated, too, as there is a clear bimodal distribution to this variable as seen from the lower plot in Figure 12.1; the two peaks correspond to low and high frequency auxiliaries. We will therefore use this variable in what follows.

We will now start to build the appropriate model for interpretation. As usual, four models are constructed and compared. The first model contains the binned frequency predictor along with varying intercepts for verb lemma and text.

freqaux.ri<-glmer(RealizationOfPAC~(1|LemmaParticiple)
+(1|Text)+BinnedFreqAux,pac,family=binomial)

The next three models contain varying slopes in addition to varying intercepts. The second model therefore contains the frequency predictor along with varying intercepts for verb lemma and text, as well as varying slopes for the verb’s lemma.

freqaux.rs.lemma<-glmer(RealizationOfPAC~(BinnedFreqAux|LemmaParticiple)
+(1|Text)+BinnedFreqAux,pac,family=binomial)

The third model contains the predictor along with varying intercepts for verb lemma and text, as well as varying slopes for text.

freqaux.rs.text<-glmer(RealizationOfPAC~(1|LemmaParticiple)
+(BinnedFreqAux|Text)+BinnedFreqAux,pac,family=binomial)

Finally, the fourth models contains the predictor with varying intercepts and slopes for both verb lemma and text.

freqaux.rs.both<-glmer(RealizationOfPAC~(BinnedFreqAux|LemmaParticiple)
+(BinnedFreqAux|Text)+BinnedFreqAux,pac,family=binomial)

The deviance ($-2LL$) and $AIC$ values (along with related statistics) are reported in Table 12.13. When each of the three more complex models (freqaux.rs.lemma, freqaux.rs.text, freqaux.rs.both) are compared with the more simple varying intercepts model (freqpart.ri), all of them are significantly better. However, which one of the more complex models should be selected for further appraisal? We will therefore compare the most complex model freqaux.rs.both with the nested models freqaux.rs.lemma and freqaux.rs.text in turn, using the information in Table 12.13. When varying slopes for both the multilevel variables are included, that model results in a significantly smaller deviance than when only by-text slopes are included (10.41 = 2937.13 – 2926.72, $p = 0.00548905$) and when only by-lemma slopes are included (36.24 = 2962.96 – 2926.72, $p = 1.350778e-08$). Consequently, we can conclude that the most complex model, that is the one with both varying intercepts and varying slopes, should be selected for further appraisal.

Before we can interpret the model, it is necessary to compare the selected model (freqaux.rs.both) with the baseline vcm. An ANOVA is presented in Table 12.14. This shows quite clearly that the selected model, with varying intercepts and varying slopes for both of the multilevel variables, is much more
### Table 12.13: Deviance statistics for various glmer models containing auxiliary frequency information

<table>
<thead>
<tr>
<th>Model</th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>freqaux.ri</td>
<td>4</td>
<td>2982.69</td>
<td>3005.84</td>
<td>-1487.34</td>
<td>2974.69</td>
</tr>
<tr>
<td>freqaux.rs.lemma</td>
<td>6</td>
<td>2974.95</td>
<td>3009.68</td>
<td>-1481.48</td>
<td>2962.95</td>
</tr>
<tr>
<td>freqaux.rs.text</td>
<td>6</td>
<td>2949.13</td>
<td>2983.85</td>
<td>-1468.56</td>
<td>2937.13</td>
</tr>
<tr>
<td>freqaux.rs.both</td>
<td>8</td>
<td>2942.72</td>
<td>2989.01</td>
<td>-1463.36</td>
<td>2926.72</td>
</tr>
</tbody>
</table>

### Table 12.14: ANOVA of baseline vcm and glmer containing auxiliary frequency information

<table>
<thead>
<tr>
<th>Model</th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>glmer.vcm.both</td>
<td>3</td>
<td>3139.41</td>
<td>3156.77</td>
<td>-1566.70</td>
<td>3133.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>freqaux.rs.both</td>
<td>8</td>
<td>2942.72</td>
<td>2989.01</td>
<td>-1463.36</td>
<td>2926.72</td>
<td>206.69</td>
<td>5</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Informative than the baseline vcm, because it has a lower AIC and its reduction in deviance is highly significant.

With the final model in place, we will now turn to an interpretation of the parameters, which are reported in Table 12.15. The second line of the table says that when the frequency category of the auxiliary changes from a low frequency category to a high frequency category, the logit probability of a factus est outcome increases by 1.1606 units. This is equivalent to an odds ratio of \( \exp(1.1606) = 3.191848 \). As we can see from Table 12.15, when this estimate is divided by its standard error (1.1606/0.2745), the z-statistic yields a value greater than the large sample critical value of 1.96, that is 4.228, indicating that the parameter estimate is highly significant indeed.

| Estimate          | Std. Error | z value | Pr(>|z|) |
|-------------------|------------|---------|----------|
| (Intercept)       | -0.1711    | -1.010  | 0.313    |
| BinnedFreqAuxhigh | 1.1606     | 4.228   | 2.36e-05 |

### Table 12.15: glmer of serialization ∼ auxiliary frequency

It is also worth mentioning the varying slopes here, given their overall importance in the model. Recall that the intercept refers to the grand mean of intercepts, and the slope refers to the grand mean of slopes. That is, the slope refers to the average value of the effect of frequency, taking into account by-text and by-lemma variability. If we wanted to find the likelihood of obtaining a factus est outcome in Caesar’s Gal. for the verb lemma morior ‘die’ for this particular variable, we plug in the relevant information into the formula. According to the information established by the model, the intercept adjustment for Caesar’s Gal. is 0.25577062, while the intercept adjustment for morior is 0.512240013; this means that both for Caesar’s Gal. and for morior, factus est outcomes are generally more likely, and the intercept has been underestimated. Similarly, the slope adjustment for Caesar’s Gal. is 0.5297853270, while the slope adjustment for morior is \(-1.031911226\); this means that the frequency of the auxiliary effect is stronger in Caesar’s Gal. than on average, but it has the opposite effect for morior, i.e. high frequency auxiliaries constructed with participles from morior appear to prefer est factus outcomes. This latter statistical point can be seen in impressionistic terms by printing all the observations with morior as the verb lemma: here 1 out of 3 constructions with a high frequency auxiliary are realized as factus est, but by contrast 4 out of 5 constructions with a low frequency auxiliary are realized as factus est. The effect is the reverse of what it is for all verb lemmas averaged out.

```r
> pac$Observation[pac$LemmaParticiple=='morior']
[1] mortuus.est est.mortuus mortuus.sit erat.mortuus est.mortuus mortuus.erat mortuus.erat mortui.erunt
```
In sum, higher frequency participles and auxiliaries are more likely to be serialized as factus est, while by contrast lower frequency participles and auxiliaries are more likely to be serialized as est factus. Both variables are compatible with the morphosyntactic interpretation of frequency effects, in that words of high frequency forms resist being realized in innovative forms. The factus est serialized is regarded as being the historically and more conservative form, so this would make historical sense. One would therefore assume that when both the participle and the auxiliary are of low frequency, est factus should be preferred, whereas when they are both of high frequency, factus est should be preferred. This can be seen in Table 12.16, which shows the rate of PAC serializations cross-classified for whether the participle and the auxiliary are both of low frequency, of mixed frequency, and of high frequency.

<table>
<thead>
<tr>
<th></th>
<th>est factus (%)</th>
<th>factus est (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>both low</td>
<td>528 (53.82)</td>
<td>453 (46.18)</td>
</tr>
<tr>
<td>mixture</td>
<td>338 (33.20)</td>
<td>680 (66.80)</td>
</tr>
<tr>
<td>both high</td>
<td>93 (22.68)</td>
<td>317 (77.32)</td>
</tr>
</tbody>
</table>

Table 12.16: Cross-tabulation of PAC serialization and combined participle and auxiliary frequency

This table establishes that when both the auxiliary and the participle are of low frequency, the innovative variant has the greater share (53.82%), e.g. est proficiscendum ‘he was obliged to depart’; that when the auxiliary and participle are of a mixed frequency, the share is approximately as it is on average; and that when the participle and the auxiliary are of high frequency, the rate of the conservative word order is the strongest as it has the larger share (77.32%), e.g. factum est ‘it was done’.5

However, the results in this chapter and the data in Table 12.16 can be interpreted in a different way. Specifically, we can focus on the psycholinguistic and cognitive aspects of the frequency effects. Going on the participle frequency effect, we might assume that the participle is put earlier, and hence the PAC is realized as factus est, because it is straightforward to process. This would correspond to the well-established familiar > unfamiliar hierarchy. However, when the participle is of a low frequency, the auxiliary is put first, as it is put into service to ‘buy time’ for accessing and articulating the participle. However, that would essentially make incorrect predictions about what we observe for auxiliary frequency, because it is placed first even though it is of low frequency. Nevertheless, the effect could be stated in broader rather than specific psycholinguistic terms: it seems that factus est is used when processing and accessibility constraints are low (i.e. when the participle and the auxiliary are of high frequency) and est factus is used when processing costs are quite high (i.e. when the participle and the auxiliary are of low frequency).

It was also found that statistical modelling is not particular sensitive to gradient frequency effects, as established for both variables. However, a point of difficulty concerns the binning of auxiliary frequency. The ctree() information yielded two bins of low frequency (≤ 25227) and those of high frequency (> 25227). As it turns out, those of high frequency are in fact all of the word form est, and the results are consequently very similar to those we obtained for auxiliary clisis (est is without doubt clearest clitic form, as discussed earlier) and for other information on the auxiliary’s form (e.g. number and TAM). More robust statistical procedures will therefore be required to tease apart the effects of clitichood and frequency, a point we turn to in the next part of the thesis.

5This table is statistically significant ($\chi^2 = 149.1304$, df = 2, $p < 2.2e - 16$), but I will not pursue the modelling of it in the present chapter.
12.4 Summary

In this chapter we have explored the effects of two sets of usage-based variables, namely those concerning persistence (the use of the same structure in a previous variable site, termed the *prime*, and the observation site, termed the *target*) and frequency. While no main effect of persistence was found, it was found to participate in a significant interaction with a variable tapping the similarity of the prime’s and target’s auxiliary. Frequency too was shown to have an effect, with higher frequency participles and auxiliaries favouring *factus est.*
Part IV

Multivariate Modelling
Chapter 13

Methodological Preliminaries

13.1 Introduction

In Part III we enumerated the information space, Ω, the many variables that can potentially bear on morphosyntactic variability. They were motivated theoretically and empirically, and we witnessed how they exert influence on PAC serialization in isolation through a series of bivariate multilevel models. However, the analysis is by no means complete, as it is not yet multivariate. That is, we have not considered how all the predictors in the information space contribute to the alternation jointly. Two predictors \( x_1 \) and \( x_2 \) may both be of influence when entered individually into separate models, but it may be that the information \( x_2 \) contributes is completely included within the information of \( x_1 \), thereby rendering the effect of \( x_2 \) epiphenomenal. Furthermore, a predictor \( x_3 \) may be uninfluential on its own (that is, as a main effect), but it may participate in a significant interaction with one or more predictor term(s). To allow these effects to become clear, multivariate modelling is required. This chapter discusses the multivariate technique of random forests to be used to assess the joint influence of all the predictors considered thus far. The dative alternation of Bresnan et al. (2007) will illustrate the methodology.

13.2 Motivating random forests

One possibility would simply be to include all the predictors in a multiple \texttt{glmer}. However, there are problems with such an approach. First, the \texttt{pac} dataset of \( n = 2409 \) observations comprises in all \( p = 33 \) covariates, some of which are factors with many levels. The number of observations, while not excessively small compared to the number of predictors, may not be sufficient when we consider all simple and higher-level interactions. For all simple interactions, we add an additional 528 terms. Considering all three-way interactions between the predictors, we add an additional 5456 terms. If we examine even higher-order interactions, such as four-way interactions, the number of additional terms used as regressors rises to 40920. Clearly, the model matrix, particularly in the latter case, results in a small \( n \times p \) problem. In \texttt{glmer}s, interactions need to be pre-specified on a theoretical basis, rather than allowing for the data to reveal its own structure.

The second problem characterizing the dataset is that 7 variables have missing observations on them. One could omit these variables, but it might result in predictive loss as those variables might be influential. One thinks here of the definiteness predictor, which showed a bivariately important association to the response in the \texttt{glmer} but is recorded on only 1511 data points, that is, those where the preverbal constituent is an NP or PP. Alternatively, one might omit all the rows containing missing values on them.

---

\(^1\)The term “multivariate” is used in an informal sense here to refer to ‘many variables considered as predictors jointly’. Strictly speaking, multivariate modelling concerns a \( y \) matrix with more than one response.

\(^2\)One can also impute missing data if so desired (see e.g. Gelman & Hill, 2007).
but again those observations might include informative information on other predictors. Note that if this
is done with the PAC dataset, 1314/2409 (54.55%) observations are lost.

Third, the PAC has several collinear predictors, that is they are related to each other in a bivariately
statistically significant way. We have already seen that (multi)collinearity is a problem in regression,
for instance, when including a categorical variable, k − 1 levels are included to avoid perfect collinearity.
To exemplify, the predictor of potential auxiliary clitichood, AuxiliaryClitichood, and that for the
frequency of the auxiliary, FreqAux, are strongly collinear: a standard GLM where AuxiliaryClitichood
is the response and FreqAux is the single predictor reveals that the predictor explains 93% (Nagelkerke
$R^2$) of the variance in the response. This connection is obvious, frequency results in phonetic reduction
leading to clisis. A variance inflation factor ($VIF$) score for a standard GLM fitted with these two as
predictors and the PAC variant as the response reveals that the variance for the each of the predictors
is inflated by 7.83. Ideally, $VIF$ should be below 4 (but views vary). Again, bottom-up, data-driven
exploration is difficult using GLMERS.

It is therefore unclear how such problems can be straightforwardly overcome with GLMER modelling,
that being unwise to throw “everything but the kitchen sink” into the same regression mode (Tagliamonte,
2012: 136). Nonetheless, such statistical problems do not beset the multivariate technique to be exploited
in this analysis — random forests — whose features we now turn to.

13.3 Conditional inference trees

Random forests are based on conditional inference trees (Hothorn et al., 2006b), a non-parametric unbi-
ased recursive binary partitioning method that draws on and improves upon standard classification and
regression trees, such as CART (Breiman et al., 1984) and C4.5 (Quinlan, 1993). Conditional inference
trees establish what is influential in classifying a response variable using the following algorithm.

To begin with, the algorithm takes the full dataset containing the response variable and the predictor
variables, and asks whether any of predictor variables are associated with the response. In statistical
terms, it tests the global null hypothesis that any of the predictors and the response are independent
of each other. The algorithm works through each of the predictors in turn, asking whether they are
important, which is assessed on the basis of test statistics or Bonferroni corrected p-values that correct
for multiple comparisons. If none of the predictors are found to be associated with the response — that
is, if the global null hypothesis cannot be rejected at the specified $\alpha$ level — the algorithm terminates.
The tree resulting from it has one terminal (or “leaf” node). However, if one of the predictors is found
be important in this first step, it is selected. If there are competing influential predictors, the most
important predictor based on the p-value is selected — that is, the one that is most strongly associated
with the response $y$. At this stage, two new subsets of data are created: one subset (SUBSET 1) contains
those observations with the values (or levels) of the selected predictor that are more likely to result in
one of the levels of the response variable (e.g. factus est), and the other subset (SUBSET 2) contains
those observations with the values (or levels) of the selected predictor that are more likely to result in
the other the level of the response variable (e.g. est factus). The full dataset has therefore been split
(partitioned) into two (binary) new subsets of data. The tree resulting from it therefore has two terminal
nodes. The algorithm then starts over again with these subsets of data in turn. It begins with SUBSET
1, and works through each of the predictors in turn, asking whether they are important, which is again
assessed on the basis of a p-value. If none of the predictors are important, the algorithm will proceed to
SUBSET 2 and so on before eventually terminating when no further significant splits are found. If one of
the predictors is found to be important for the data in SUBSET 1, it is selected. If there are competing
influential predictors, the most important one based on the p-value is selected.

At the next stage, two further subsets of data are created: one subset (SUBSET 3) contains those

---

3The exact way the p-value is computed will not concern us here.
observations with the values (or levels) of the selected predictor that are more likely to result in one of the levels of the response variable (e.g., \textit{factus est}), and the other subset (subset 4) contains those observations with the values/levels of the selected predictor that are more likely to result in the other level of the response variable (e.g., \textit{est factus}).

The algorithm then starts over again with these subsets of data in turn. The algorithm continues until there are no more significant binary splits to be made. In this respect, the algorithm partitions the dataset recursively into smaller non-overlapping subsets of data. See Figure 13.1 for a diagrammatic representation of the tree at this stage.

Let us concretize this procedure by going through the familiar case study of the dative alternation (Bresnan et al., 2007), the dataset of which is available in the \texttt{languageR} package for R. The response variable is an NP outcome (Mary gave John the book) or a PP outcome (Mary gave the book to John). As predictors, let us consider the following: a predictor termed \texttt{AnimacyOfRec}, which has two levels, animate vs. inanimate, identifying the animacy of the recipient; a predictor termed \texttt{DefinOfRec}, which has two levels indefinite vs. definite, identifying the definiteness of the recipient; and a predictor termed \texttt{AccessOfRec}, with the levels given vs. accessible vs. new, identifying the accessibility of the recipient in the discourse. Of course, many more variables are important and included in the dataset, but for illustration purposes only these three are chosen. The conditional inference tree fitted to the data uses the \texttt{ctree()} function in the \texttt{party} package in R (Hothorn et al., 2006a):

\begin{verbatim}
dative.ctree<-ctree(RealizationOfRecipient~AnimacyOfRec+DefinOfRec+AccessOfRec,dative)
\end{verbatim}

The resulting conditional inference tree is given in Figure 13.2.

Let us now walk through the tree detailing the procedure involved. In the first step, the full dataset is taken, containing $n = 3262$ observations, of which 2414 (73.98%) are realized as an NP recipient and 849 as a PP recipient. The algorithm first tests the global null hypothesis that any of the three candidate predictor variables are significantly associated with the response. It turns out this is the case, otherwise we would have ended up with a single terminal node with 73.98% shaded light grey and 26.02% shaded dark grey. On the basis of a $p$-value (or test statistic) for the null hypothesis of each predictor variable and the response, the predictor is chosen which exhibits the strongest association with the response. In this case, it turns out to be \texttt{AccessOfRec}. Two new subsets of data are then created. The data are split according to the levels favouring one or the other variant. In this case, \texttt{AccessOfRec} is a multiclass factor with three levels: given recipients on the whole associate with the NP realization, while accessible and new recipients associate on the whole with the PP realization. The 2303 observations with given recipients are therefore isolated from the rest of the observations containing accessible and new recipients. The observations with given recipients form one subset of data (node 2 in Figure 13.2) and the observations with accessible/new recipients form a second subset of data (node 5 in Figure 13.2).

The algorithm then works left-to-right recursively. Specifically, it then works through the candidate predictors in the subset termed node 2, and it determines that \texttt{AnimacyOfRec} is significantly influential.
This subset of data is then partitioned in two according to the factor levels that more likely favour NP vs. PP recipients. We can see that, on the whole, animate recipients are proportionally more likely to result in an NP realization than are inanimate recipients. Thus, a subset of data containing observations with inanimate recipients \((n = 132)\) is isolated from those observations containing animate recipients \((n = 2170)\), forming nodes 3 and 4, respectively.

The subset of data in node 3 is then tested for whether the remaining predictor DefinOfRec is significantly associated with the response. It is not, so we thus have a terminal node here. The algorithm proceeds left-to-right to the newly created subset of data in node 4, to determine whether the remaining predictor DefinOfRec is significantly associated with the response in this subset as well. Again, it is not, so we end up with a terminal node, and proceed left-to-right to node 5.

The algorithm works through node 5 and evaluates the global null hypothesis of independence between the response and any of the covariates. There again is a significant association with regard to AccessOfRec (recall that this predictor is still available for binary splitting). A binary partition is performed between observations containing the level new and those observations containing the level accessible. The subset of data for the former, node 6, contains 346 observations, while that for the latter, node 7, contains 615 observations.

The algorithm then works through the subset in node 6 to determine whether the predictors of animacy and/or definiteness exhibit any significant association to the response. It appears that the global null hypothesis of independence cannot be rejected, so the algorithm moves to the subset of data in node 7. Here, the global null hypothesis can be rejected, specifically with AnimacyOfRec carrying a strongest association with the response variable. A binary partition is initiated between observations recorded on the level animate \((n = 549)\) and those observations recorded on the level inanimate \((n = 66)\).

There are no more significant splits to be made, so the algorithm terminates at this point. We therefore end up with a conditional inference tree of 5 terminal nodes, which characterizes the dataset in terms of the relationship between the response and a rather small information space of three variables. In sum, we can see that accessibility and animacy are important, while definiteness is not, at least here.

Before we move on, there are two points to note. First, trees show the hierarchy of important predictors: here, accessibility > animacy > (definiteness). A second point concerns the fact that they demonstrate interaction effects. For instance, there is an interaction between accessibility and animacy in Figure 13.2. While there is no effect of animacy for new recipients, there is for given and accessible recipients, whereby inanimates favour the PP realization. Such an “asymmetric” interaction in a GLM or GLMER would have to be pre-specified in advance on theoretical bases, but here the interactions are identified through the recursivity method in a highly informative, exploratory, data-driven way. Finally, it is worth noting that such a method is unbiased because, while traditional tree-based methods such as CART typically favour numeric predictors or categorical predictors with many levels, conditional inferences trees as implemented by the ctree() function in R do not.

### 13.4 Growing a random forest of trees

While a single conditional tree has useful properties, such as demonstrating interactions, there are problems with them. For instance, the recursive binary partitioning algorithm heavily depends on the learning data sample. Minor changes in the data can result in different predictors being selected and predictors being differently partitioned. This is particularly the case when the dataset has relatively few observations and the covariate matrix is vast. As such, this may result in overfitting and the single conditional inference tree may not be generalizable globally, and they are only locally optimal.

Such problems can be remedied by means of random forests. While a single conditional inference tree is just that, a random forest is a whole forest of such trees, which is a type of ensemble method (Breiman,
Figure 13.2: Conditional inference tree for the English dative alternation containing three predictors: the animacy, definiteness, and accessibility of the recipient.
13.4. GROWING A RANDOM FOREST OF TREES

It is random in several respects. First, different subsets of observations are bootstrap sampled with replacement. Individual conditional inference trees are then constructed for each different subset of observations. In the `cforest()` function in R, available in the `party` package, the default number of bootstrapped samples is 500. This means that for a forest grown with the default number of bootstrapped samples, 500 conditional inference trees belong in that forest. Second, in a standard conditional inference tree, as demonstrated above, all the available predictor variables are evaluated at each potential splitting stage for the strength of their association with the response variable; the predictor with the strongest association is selected for splitting. In random forests, by contrast, it is possible to evaluate all predictors at each splitting stage, in which case it is known as “bagging”, but it is also adds a further dimension of randomness to restrict the number of candidate variables at each splitting stage. This allows the forest of trees to be more diverse, enhancing the generalizability of the results for influential predictors. The benefits of this can be summed up thus:

In addition to the smoothing of hard decision boundaries, the random selection of splitting variables in random forests allows predictor variables, that were otherwise outplayed by a stronger competitor, to enter the ensemble: If the stronger competitor cannot be selected, a new variable has a chance to be included in the model and may reveal interaction effects with other variables that otherwise would have been missed (Strobl et al., 2009b).

One final dimension of randomness is the forest’s seed, which can be any number. Therefore, in order to establish whether a model is robust or not, it is advisable to run at least two models on different seeds. If different results are obtained, the number of conditional inference trees that are grown should be increased.

In order to establish prediction accuracy and assess the contribution of each predictor considered, random forests make use of “in-bag” and “out-of-bag” observations. For each conditional inference tree in the forest, a bootstrap sample of the dataset is taken. The observations which form this bootstrap sample are known as the in-bag observations for that tree, while the observations which are not used are known as the out-of-bag observations for that tree.

In terms of prediction accuracy — to what extent the predicted values match up with actually observed values — a method of “voting” is used. Each out-of-bag observation \( y_{i,\text{OOB}} \) is passed through the tree informed by the in-bag observations. Given the levels/values of the predictor variable, each tree then casts a vote as to which of the levels of the response variable it thinks it belongs to. The overall result is an \( n \times n_{\text{tree}} \) matrix of votes for each observation in the dataset, where \( n \) is the number of observations and \( n_{\text{tree}} \) is the number of trees in the ensemble.

Simplifying somewhat, let us assume that we have constructed a very small forest with 12 trees with in-bag observations. A single observation, which is coded for the response and a suite of predictor variables, is ‘in-bag’ in 5 of the trees and ‘out-of-bag’ in 7 of the trees. It is then passed through the 7 trees in which the observation was not used to inform the tree. Each tree then casts a vote as to its likely classification, \( a \) or \( b \). Let us assume that the votes associated with this observation are as follows: \( [a, a, a, b, a, a, b] \). The votes are then averaged out of all the trees used, i.e. \( a \) obtains 5/7 (71.42%) votes and \( b \) obtains 2/7 (28.57%) votes. Therefore, for this single observation, the trees on the whole predict that \( a \) is the more likely candidate. The predictive accuracy can be assessed either through computing the proportion of correctly predicted votes or through the index of concordance \( C \) (or the \( C \)-statistic) which uses predicted probabilities for a given response observation. The technical details of this \( C \)-statistic, which is based on the Receiver Operator Curve (ROC), need not concern us. What is important, however, is that \( C \) quantifies the model’s ability to discriminate between one level of the response (e.g. \( a \), factus est, NP-dative, . . . ) and the other level of the response (e.g. \( b \), est factus, PP-dative, . . . ). Hosmer & Lemeshow (2000: 160–164) suggest that \( C \)-values equal to 0.5 indicate no discrimination, those \( 0.7 \geq C < 0.8 \)
indicate acceptable discrimination, those \(0.8 \leq C < 0.9\) indicate excellent discrimination, and those greater than 0.9 indicate outstanding discrimination.

In terms of establishing which predictors matter, while it is possible to inspect an individual tree, it is not possible to inspect an “average tree” in the ensemble, as Strobl et al. (2009b) phrase it. Instead, we focus on establishing the permutation accuracy importance of a given predictor. In order to compute this, the original values/levels of a given predictor \(x\) are randomly “permuted”. The effect of this is to break the observed association between the predictor and the response. What does this mean? Returning to the dative dataset, Table 13.1 shows 10 observations, along with the observed animacy predictor and the permuted animacy predictor. Under the column entitled ‘Observed Animacy’, we can see that the level \texttt{inanimate} always results in a \texttt{pp}-dative, whereas the level \texttt{animate} results in an \texttt{np}-dative 5/6 times. If the original values are permuted, we obtain something such as in the column entitled ‘Permuted Animacy’. This shows that now two inanimate observations yield \texttt{np}-outcomes and the remaining two inanimate observations yield \texttt{pp}-outcomes; three animate observations yield \texttt{np}-datives, and the remaining three animate observations yield \texttt{pp}-datives. The association between animacy and the response variable is thereby destroyed.

Variable importance is then determined as the difference in prediction accuracy with the original, unpermuted predictor and with the permuted predictor, as averaged across all the trees in the ensemble. Clearly, if a predictor is influential with respect to the response, then when it is permuted, the predictive accuracy of the model will decrease. By contrast, if the difference in prediction accuracy remains the same, or decreases or increases very slightly, after permuting a given predictor, then that predictor is worthless.

In order to interpret permutation variable importance, as implemented in the \texttt{varimp()} function in R’s \texttt{party} package, Strobl et al. (2009b: 21) recommend the following:

All variables whose importance is negative, zero or has a small positive value that lies in the same range as the negative values, can be excluded from further exploration. The rationale for this rule of thumb is that the importance of irrelevant variables varies randomly around zero. Therefore positive variation of an amplitude comparable to that of negative variation does not indicate an informative predictor variable, while positive values that exceed this range may indicate that a predictor variable is informative.

<table>
<thead>
<tr>
<th>Response</th>
<th>Observed Animacy</th>
<th>Permuted Animacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>animate</td>
<td>inanimate</td>
</tr>
<tr>
<td>PP</td>
<td>inanimate</td>
<td>inanimate</td>
</tr>
<tr>
<td>NP</td>
<td>-animate</td>
<td>animate</td>
</tr>
<tr>
<td>NP</td>
<td>animate</td>
<td>inanimate</td>
</tr>
<tr>
<td>PP</td>
<td>inanimate</td>
<td>animate</td>
</tr>
<tr>
<td>PP</td>
<td>animate</td>
<td>inanimate</td>
</tr>
<tr>
<td>NP</td>
<td>animate</td>
<td>animate</td>
</tr>
<tr>
<td>NP</td>
<td>animate</td>
<td>animate</td>
</tr>
<tr>
<td>PP</td>
<td>inanimate</td>
<td>animate</td>
</tr>
</tbody>
</table>

Table 13.1: Observed predictor and permuted predictor of animacy based on Tagliamonte & Baayen (2012)

Strobl et al. (2008, 2009a) demonstrate that the permutation variable importance measure, as established by Breiman (2001), is problematic on datasets with highly correlated data, in which the correlated predictors cluster high in variable importance, when only one of them might actually be important. To avoid this issue, they suggest computing the conditional variable importance of a predictor. Such a measure has been used in the language variationist research of Tagliamonte & Baayen (2012). However, this advantage aside, it has a number of disadvantages. First, it is a computationally intensive procedure,
which requires hours (or days) of processing time. For example, Tagliamonte & Baayen (2012) observe that “[e]stimating the conditional variable importance for the full data set required approximately 8 hours of processing time on a state-of-the-art CPU”. Second, there is as yet no way to factor in missing values (Carolin Strobl, p.c.), which characterize the PAC dataset. As such, conditional variable importance would have to be conducted on only those observations with no missing data recorded on them, reducing thereby the amount of data random forests has to select from and causing observations to be omitted which might be of use on other predictor variables.

There may be one further justification for using unconditional permutation accuracy importance. Aside from it being computationally faster and able to handle missing values on the predictors, Strobl et al. (2008) demonstrate that with increasing values of the tuning parameter $m$ — that is, the number of preselected predictors available for evaluation at each split in the conditional inference tree — the unconditional importance resembles the behaviour of the conditional importance. The relevance of this will be discussed in more depth when we consider the application of random forests to the PAC alternation in the next chapter.

13.5 Application of random forests in language research

It is worth pointing out that the application of conditional inference trees and random forests, as opposed to mainstream statistical methodologies, in language research is new and underappreciated. While they are now gaining ground in disciplines of genetics, epidemiology, and psychology — all of which involve the analysis of high-dimensional datasets — there are relatively few studies that employ such methods. For instance, they have been used to evaluate variability in the positioning of concessive clauses (Wiechmann & Kerz, 2013), featural similarities in contrastive linguistics (Wiechmann, 2011), /s/-voicing in Quito Spanish (Strycharczuk et al., 2014), variability in English dative and genitive alternations (Shih & Grafmiller, 2011), sociogrammatical variation between was/were in the English of York (Tagliamonte & Baayen, 2012), and predictors of phonetic reduction (Dilts, 2013). There is a growing need for such models in language research, though, with datasets become ever more complex and multidimensional. Because of this, and comparing them with GLMs, GLMERS, and single conditional regression trees, Tagliamonte & Baayen (2012) diagnose random forests as “the ideal panacea for the thorniest problems of variation analysis”. Indeed, the efficacy of random forests has been characterized by Tutz (2012: 458, 459) as being “one of the most efficient classifiers that have been proposed in recent years”, who points out further that “in many studies random forests were shown to be among the best predictors”. Given the relative dearth of linguistic applications, it is one of the aims of the present study to demonstrate how they can be of benefit to understanding the complex structure of historical language datasets. In the next chapter, a random forests model of the Latin PAC alternation is to be constructed using the methods and steps identified in this chapter.
Chapter 14

Random Forests Model of the PAC Alternation

14.1 Introduction

With the statistical methodology introduced in depth in the previous chapter, we turn in this chapter to present a random forests model of the Latin PAC alternation. Section 14.2 outlines the procedure specific to this dataset. Section 14.3 presents and discusses the results in detail. In Section 14.3.1 I discuss the overall predictive and discriminatory accuracy of the random forests model; in Section 14.3.2 I compare the variable importance ranking of the random forest model with that of individual GLMs for each of the predictors.

14.2 Procedure

In this section I discuss the random forests procedure specific to the PAC dataset. As for the modelling, the R party package (Hothorn et al., 2006a) is exploited. The cforest() function is used to grow the forest, and the varimp() function is used to assess the relative importance of the predictors.

Thirty-two of the predictors introduced in Part III are included in the random forests model, including the moderator variables for persistence (previous lemma similarity and previous auxiliary similarity) and the random effect factor of text. However, given the excessive number of levels for the lemma \( k = 604 \), it was decided to omit this variable to facilitate computation of the forest and the permutation accuracy scores. Even if those factor levels which did not meet a minimum criterion of, for example, 6 observations are collapsed to form a predictor with \( k = 94 \) levels, and the cforest() algorithm is run, the inclusion of lemma renders the procedure computationally intractible. The algorithm meets with convergence problems. Frequency predictors are here entered as metric covariates to allow for more sensitive interactions along their continuous values.

As mentioned in the previous chapter, because of the random nature of the algorithm, it is necessary to modify the tuning parameter ntree in the cforest() function in order to ensure stable variable importance rankings when run on different random seeds (Strobl et al., 2009a). To get an idea of the number of trees that result in a fairly stable random forest for this dataset with 32 predictors, random forests were therefore grown initially on ntree = 100 on two different seeds, which was then incrementally increased by 250 trees up to an ntree value of 2500. Throughout the forest runs, the tuning parameter mtry was set to the square root of the number of inputs \( \sqrt{p} = \sqrt{32} \approx 6 \), as is generally recommended (e.g. Strobl et al., 2009a). The most optimal size of ntree resulting in an across-seed stable forest was 2500. However, this falls somewhat short of a perfect match, as only 19 of the 32 predictors are in exactly the same position relative to each other in terms of their variable importance ranking. Increasing the

---

1 An alternative randomForests is also available, but this uses classification and regression trees rather than conditional inference trees based on unbiased recursive binary partitioning.
value of \texttt{ntree} even further simply becomes computationally unviable given the hardware available and the feature space under consideration. For simplicity, then, I will simply report the results of the most optimal \texttt{ntree} value.\footnote{To compensate for this, one possibility would be to take the average results of (e.g.) 50 forests grown with the computationally optimal value of \texttt{ntree} for the dataset, but given time constraints, it was decided not to do this. The random seed of this forest was \texttt{set.seed(4134)}.}

### 14.3 Results and discussion

#### 14.3.1 Prediction accuracy

We turn first to discriminatory power and predictive accuracy of the random forest. The discriminatory index $C$-statistic for the random forest model ($\texttt{seed} = 4134$, $\texttt{mtry} = 6$, $\texttt{ntree} = 2500$) — when predicted probabilities are used from the entire sample — is 0.9202, which is “outstanding” discriminatory power in the terms of Hosmer & Lemeshow (2000). When predicted probabilities are used from only the out-of-bag observations, $C$ falls to 0.8025, which is just in “excellent” territory in the terms of Hosmer & Lemeshow (2000).

The percentage of correct predictions across the entire sample is 84.06%, while the predictive accuracy using only the out-of-bag observations — the more conservative measure — is 72.93%. The confusion matrices for the cross-sample and out-of-bag observations are given in Table 14.1 and Table 14.2, respectively. Jittered separation plots for the same are given in Figure 14.1. The latter shows the concentration of predictions effectively. What these show is that the classifier is far more able to classify instances of \textit{factus est}, than \textit{est factus}, that is there are more uncertain predictions with regard to the later. The reason behind this may be that, first, there are more instances of \textit{factus est} overall, and, second, we have discovered predictors and values/levels in this study that are strongly associated with \textit{factus est}. Recall from the \texttt{glmer}s in the previous part of the thesis that certain levels of the predictors concerning auxiliary clitichood (e.g. the level \textit{clitic}) and preverbal constituent (e.g. the level \textit{nothing}) very strongly associate with \textit{factus est}. By contrast, there is little in the information space that unanimously associates with the \textit{est factus} serialization, except (e.g.) a preverbal constituent that is the negative adverb \textit{non} and concessive clauses. Further research will therefore be required to discover what is important to \textit{est factus}, but \textit{factus est} poses no such challenge.

Predictions in terms of the predicted response class and the predicted class probabilities for some randomly sampled observations are reported in Table 14.3. These are entire sample predictions. In sum, the random forest model is a capable and robust learner. A competing classifier, a bagged forest model, where \texttt{mtry} is not random but set to equal the number of inputs, exhibits similar predictive performance across the entire sample at 84.18% and for out-of-bag observations at 72.44%. In comparison to a single conditional tree grown with all predictors in it, random forests is superior. The tree’s predictive

<table>
<thead>
<tr>
<th>Observed Serialization</th>
<th>Predicted Serialization</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{est factus}</td>
<td>685</td>
</tr>
<tr>
<td>\textit{factus est}</td>
<td>274</td>
</tr>
</tbody>
</table>

Table 14.1: Random forest confusion matrix for entire sample

<table>
<thead>
<tr>
<th>Observed Serialization</th>
<th>Predicted Serialization</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{est factus}</td>
<td>542</td>
</tr>
<tr>
<td>\textit{factus est}</td>
<td>417</td>
</tr>
</tbody>
</table>

Table 14.2: Random forest confusion matrix for out-of-bag observations

of \textit{factus est}, than \textit{est factus}, that is there are more uncertain predictions with regard to the later. The reason behind this may be that, first, there are more instances of \textit{factus est} overall, and, second, we have discovered predictors and values/levels in this study that are strongly associated with \textit{factus est}. Recall from the \texttt{glmer}s in the previous part of the thesis that certain levels of the predictors concerning auxiliary clitichood (e.g. the level \textit{clitic}) and preverbal constituent (e.g. the level \textit{nothing}) very strongly associate with \textit{factus est}. By contrast, there is little in the information space that unanimously associates with the \textit{est factus} serialization, except (e.g.) a preverbal constituent that is the negative adverb \textit{non} and concessive clauses. Further research will therefore be required to discover what is important to \textit{est factus}, but \textit{factus est} poses no such challenge.

Predictions in terms of the predicted response class and the predicted class probabilities for some randomly sampled observations are reported in Table 14.3. These are entire sample predictions.

In sum, the random forest model is a capable and robust learner. A competing classifier, a bagged forest model, where \texttt{mtry} is not random but set to equal the number of inputs, exhibits similar predictive performance across the entire sample at 84.18% and for out-of-bag observations at 72.44%. In comparison to a single conditional tree grown with all predictors in it, random forests is superior. The tree’s predictive
Figure 14.1: Separation plots for entire sample (top) and out-of-bag observations (bottom).
14.3. RESULTS AND DISCUSSION

<table>
<thead>
<tr>
<th>Observation</th>
<th>Reference</th>
<th>y</th>
<th>$\hat{y}$</th>
<th>$p(y = 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>erat adiectum</td>
<td>B. Alex. 28</td>
<td>0</td>
<td>0</td>
<td>0.15</td>
</tr>
<tr>
<td>esse commissurum</td>
<td>B. Hisp. 9</td>
<td>0</td>
<td>0</td>
<td>0.22</td>
</tr>
<tr>
<td>erat adhortatus</td>
<td>Caes. Civ. 1. 34</td>
<td>0</td>
<td>0</td>
<td>0.37</td>
</tr>
<tr>
<td>pugnatum est</td>
<td>Caes. Gal. 3. 21</td>
<td>1</td>
<td>1</td>
<td>0.98</td>
</tr>
<tr>
<td>fuerat usus</td>
<td>Caes. Gal. 5. 25</td>
<td>0</td>
<td>0</td>
<td>0.45</td>
</tr>
<tr>
<td>conturbatus esse</td>
<td>Cic. Phil. 2. 36</td>
<td>1</td>
<td>1</td>
<td>0.73</td>
</tr>
<tr>
<td>nactus est</td>
<td>Cic. Phil. 3. 31</td>
<td>1</td>
<td>1</td>
<td>0.76</td>
</tr>
<tr>
<td>pollicitus esset</td>
<td>Cic. Phil. 13. 50</td>
<td>1</td>
<td>1</td>
<td>0.70</td>
</tr>
<tr>
<td>erant oppressi</td>
<td>Hirt. 8. 24</td>
<td>0</td>
<td>0</td>
<td>0.28</td>
</tr>
<tr>
<td>oppositum esse</td>
<td>Nep. Them. 7</td>
<td>1</td>
<td>1</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Table 14.3: Randomly sampled examples of observed values and predicted values. Here, $y/\hat{y}$ is in terms of $factus est$, such that $1 = factus est$ and $0 = est factus$

accuracy is 69.12%, and has a $C$-index of 0.7524. However, conditional inference trees are still of interest, and will be utilized further below.

14.3.2 Variable importance

Permutation accuracy measures of variable importance show the importance of a variable, regardless of whether it has a main effect or interactive effect with another predictor. In order to determine whether a predictor has an effect on its own, participates in a significant interaction with another predictor, or is completely worthless can be identified by comparing the position of a predictor on the variable importance hierarchy of the random forest and its variable importance in a glm. In Figure 14.2 two plots are shown. The plot on the top shows the variable importance ranking for the random forest. The plot on the bottom shows the variable importance hierarchy resulting from comparisons of thirty-two individual standard GLMs. The latter was calculated by the following formula:

$$VI_{GLM,p} = AIC_{GLM, null} - AIC_{GLM,p}$$  \hspace{1cm} (14.1)

where $VI_{GLM,p}$ is the variable importance score for some predictor $p$, $AIC_{GLM, null}$ is the Akaike Information Criterion ($AIC$) (Akaike, 1974) for the null model specified for just the intercept ($\sim 1$) and $AIC_{GLM,p}$ is the $AIC$ for the model containing some predictor $p$. The difference in $AIC$ between the null model and the model containing some predictor indicates importance in that, for important predictors, the $AIC$ will be closer to zero and will show a larger positive $VI$, while, for uninfluential predictors, the $AIC$ will be closer to that of the null model or even be higher, thereby exhibiting a small or negative $VI$.

Although the two measures are not exactly comparable, we will nonetheless use Strobl et al.’s (2009b) rough rule of thumb for identifying potentially influential and non-influential predictors, by taking those variables with values greater than the absolute value of the lowest variable importance score to be of some statistical informativeness.

Going on this admittedly fairly coarse guiding principle, there are three variables that are deemed inconsequential by the random forest model and the glm — person, impersonality, and the previous lemma similarity (the latter only being included as a potential moderator variable). These three predictors were shown to play little of a role in the glmer analyses presented earlier. It is, however, striking that person is not important, given its influence elsewhere in grammatical variation (recall the discussion in Section 10.10). As such, the modeller of pac variability may discard these variables from further consideration.

In sum, 23 predictors are established as being potentially useful in both the random forest analysis and in the standard glm. These predictors span the many subsets of the information space considered here.

---

3I decided not to use the glmer $AIC$s here to make comparison easier. Therefore, the variable Text which was treated as a multilevel/random effect in Part III was included as a fixed effect input here.
Figure 14.2: Random forests and GLM variable importance rankings for the predictors considered in this study.
To give a couple of examples, the multilevel variable of text is influential, which is to be expected, given that idiolectal variation appears in the majority of variationist datasets (see e.g. Tagliamonte & Baayen 2012). Both the predictors pertaining to clisis — the auxiliary’s potential clitic status and the position of the PAC in what we have assumed to be the phonological phrase — are of especial influence, ranking very high overall. Indeed, PositionOfPAC ranks highest in both. The clausula predictor, PartAuxMetrics, is also important in both models.

Be that as it may, there are some differences. First, the variables of PreverbalAnimacy, PreverbalPartSegment, PreverbalTopicality, SubjectGender, PartAuxSegment, and PartAuxRhythm are of at least some importance in the random forests models, but not in the standard GLM. While their main effects are not important, they become important in conjunction with another predictor. The interactions may be low-order (the effect of the moderator depends on another predictor) or higher-order (the effect of the moderator depends on several predictors). These interactions have to be prespecified in GLMs and GLMERS, but in random forests they come to the fore in an exploratory fashion through the mathematics of the algorithm. Several examples of this can be illustrated by returning to conditional inference trees, which are excellent at showing such interaction traits. We will focus on low-order interactions here, but further analysis may reveal interesting higher-order interactions.

Preverbal animacy interacts with the variable AuxPartRhythm, the latter indicating the eurhythmicity across the auxiliary-participle border in that serialization (cf. Figure 14.3). Animacy becomes relevant only when the sequence is harmonious: a very slight effect is found, whereby factus est is favoured with np/pp denoting abstract concepts or organizations.

We noted earlier that a higher topic-worthiness of the preverbal constituent (np/pp) might have a facilitative effect on stimulating an est factus order — given that such constituents are more likely to be information structurally focused — a factor which has been claimed to give rise to est factus in the literature. We see its effect in evidence when it depends on the values/levels of other predictors (cf. Figure 14.4). First, in two texts — Nepos and Caesar’s Bellum Gallicum — the effect of topic-worthiness is enhanced. It is unclear why it should affect just these two texts, however, so this might repay further attention. Second, when the preverbal constituent is indefinite and picked up later, est factus is more likely. It seems plausible that an indefinite NP which is topicalized in the next clause would more likely be an information focus than if it was not. Thirdly, preverbal topic-worthy NPs that are not in a position to which clitics are sensitive (e.g. the level annotated other
rather than 1P, 2P) are more likely to encourage *est factus* serializations; thus, once the prosodic phonology is out of the way, topicality effects can have their way. Fourth, there is an interaction effect with number, whereby topicality is important for the serialization of PACs inflected in the singular, but not for those inflected in the plural. This variable interacts with eight further predictors in the information space (*VerbClass*, *FreqPart*, *FreqAux*, *AuxiliaryClitichood*, *AuxPartSegment*, *PartAuxMetrics*, *ConjugationalClass*, *PreverbalCategory*), where the expected effect holds only for certain values/levels of the superordinate predictor. In many of these cases (*e.g.* number, frequency), it is the value/level that is plausibly less complex in terms of access, that the topicality predictor is dependent on. That topicality participates in a number of low-level interactions when it is only on the border of relevance in the GLM’s variable importance ranking shows the effectiveness of random forests at pulling out effects that GLM/GLMERs cannot discern. Such effects warrant further attention.

As for subject gender, it serves as a moderator in nine low-level interactions, being dependent on *Text*, other grammatical predictors (*MorphologyAuxiliary*, *SubjectNumber*, *ParticipleCategory*), the usage-based variable of auxiliary frequency, and four prosodic/phonological inputs (*AuxiliaryClitichood*, *AuxPartRhythm*, *AuxPartSegment*, *PartAuxMetrics*). Some interactions it participates in are shown in Figure 14.5.

Thus, the use of fitting many conditional inference trees with bootstrapped samples of the data allow such predictors whose effects would have otherwise gone unnoticed in a GLM or GLMER to shine.

---

**Figure 14.4: Topicality interactions**
In addition to interactions noted here, it is evident from Figure 14.2 that some of the predictors switch positions in the rankings. Why, for instance, is it that Text is in sixth position in the GLM ranking, but moves into second place in the random forests ranking? The reason for this is that interaction structure has been left out of the individual GLM (each feature was entered as a main effect), whereas in the random forest it has many opportunities to enter into complex interactions with other predictors. The ranking of predictors indicates the combined importance of the predictor as a main effect and in interactions with other predictors. Evidently, Text participates in more influential interactions with respect to the response variable than the predictors that are ranked above it in the GLM hierarchy. We have already seen this to some extent in Part III through the inclusion of by-text random slopes.

14.4 Summary

This chapter has presented a random forests analysis of the PAC alternation.
Part V

Summing Up
Chapter 15

Conclusions and Additional Remarks

15.1 Summary of findings and implications

This thesis has been devoted to uncovering the predictors that bear on serial variability in Latin PACs, the first of its kind. More specifically, this construction has been exploited as a case study for capturing the probabilistic and multivariate nature of this dead language’s grammar. Hitherto, probabilistic and multivariate grammars have only really been discussed for Germanic languages on a select number of alternations. In this final chapter we will return to the battery of broad questions mentioned in the introduction:

1. To what extent are probabilistic models useful and reflective of Latin syntax variation phenomena as evidenced by the PAC alternation?
2. What are the most useful statistical models to use?
3. What types of linguistic variables influence variation?
4. What theoretical implications and explanations do the statistical models suggest?
5. What are the practical applications of probabilistic models for applied linguistics?

15.1.1 Predictors of PAC variability

This thesis has sought to confirm impressionistically mentioned effects in the previous literature, operationalize them in cross-study reliable ways, and also identify and explore the effect of new predictors, not previously mentioned with regard to the PAC alternation. It has been empirically demonstrated here that the alternation can be predicted and is shaped by a whole host of variables (termed here, the “information space”) acting in tandem. It cannot be attributable to a single predictor, nor to a linguistically coherent subgroup, such as just semantics or just phonology. Rather, different aspects of language are important, spanning the many levels of linguistic structure. In the sections that follow, the results of the various statistical models are summarized.

15.1.1.1 Multilevel structure

Two broad aspects of multilevel (random effects) structure were considered in this thesis, viz. the verb’s lemma and the text from which the observation was sampled.

Verb lemmas behave differently in how they affect PAC variability. For example, *contineo* ‘contain’ associates with *est factus*, while *coepi* ‘begin’ associates with *factus est*. They were found to be an important feature in the multilevel models discussed in Part III, and interacted with numerous predictors when entered as random slopes. Due to the vast number of levels of this random effect factor (\(k = 504\)),
verb lemma could not be included in the random forests analysis in Part IV because of computational
intractability.

Significant variability was also found with regard to text, which was here taken to stand for au-
thor/text. It was shown to be important as a varying component, and significantly interacted with many
of the predictors when included as random slopes (as evidenced by deviance scores and AICs). There
are only $k = 9$ levels to this random effect factor, so it was straightforward to include within the random
forests model. Here, it scored very highly on variable importance (2nd). It also interacted significantly in
the random forest with inputs that were not otherwise influential as main effects. Text-based variability
is a key feature of language data, and must be taken into account, at least as a control factor.

15.1.1.2 Prosodic and phonological predictors

Variation tied to prosodic and phonological features drives the alternation dramatically, and these pre-
dictors were — for the most part — found to be significantly influential in both the individual glmer
and multivariate random forests analysis. Two key predictors stand out here, namely the position of the
PAC in (what we have assumed is) the phonological phrase, and the auxiliary’s clitic potential. Moreover,
it was demonstrated in Section 9.2 that there may be two types of auxiliary clitics in Latin: phonological
phrase clitics, such as est, which avoid 1P and 2P in the intonational phrase, and intonational phrase
clitics, such as erat, which avoid 1P but seek out 2P in the intonational phrase. While the auxiliaries
do not have fully-fledged clitic status, as the effects are probabilistic, they are at least on their way to
becoming so.

Metrical, eurhythmic, and segmental phonological inputs were also found to be important,
with some conflicting findings, however. To exemplify, while the predictors PartAuxSegment,
PreverbalPartSegment, and PartAuxRhythm were deemed inconsquential by the glmer analysis (and in
terms of glm variable importance), they rise to importance in the random forests analysis. This suggests
that they interact with other predictors. Further research will be required to explore these interactions
in greater detail than can be done here.

15.1.1.3 Grammatical predictors

Eleven morphological and syntactic inputs were considered in this study: in terms of variable impor-
tance (via the random forest) these are the morphological properties of the auxiliary, the preverbal
word/constituent’s category, the participle’s category, subject number, clause type, participle prefixa-
tion, conjugational class, subject gender, deponency. There are no effects to be identified for grammatical
person, which is quite surprising given the importance of person in other respects (Section 10.10.2). Imp-
personality of the PAC was also found to have no effect, but this is not surprising, given that the literature
has regarded it as being epiphenomenal (see Section 10.5.5). Subject gender was found to be a weak but
significant predictor in the glmer analysis, where it was found to be effective when by-lemma slopes
were taken into account.

15.1.1.4 Semantic and pragmatic predictors

Two predictors of verbal semantics were examined: the eventuality aspect of the predicate and the verb
class, as approximated by its Levin (1993) classification. Both were found to be of significance in the
glmer and random forests analysis. Achievement (mori 'die') and activity (curro 'run') predicates
strongly prefer factus est serializations, whereas stative expressions (uiuo 'live') prefer est factus, in line
with previous claims.

Animacy did not have a main effect — surprising given the literature — but the random forest and
conditional inference tree analysis demonstrated that it interacted slightly with one of the phonological
predictors (Figure 14.3).
Information structure was operationalized here along three dimensions: definiteness of the preverbal NP/PP, topicality of the preverbal NP/PP (i.e. whether it is is picked up in the subsequent clause), and preverbal contrastiveness. The definiteness and contrast predictor were both significant as an effect in the glmer analysis and in the random forests. Topicality was weakly significant in its glmer model (but non-significant according to the GLM variable importance ranking), and the random forest/conditional inference tree analysis revealed interaction effects with a range of other predictors (cf. Figure 14.4).

15.1.1.5 Usage-based predictors

In the glmer models, the frequency variables for the auxiliary and the participle were binned, as they were non-linear with respect to the logit (as indicated by gams), and a conditional inference tree suggested splitting them into high vs. low frequency classes. Both were important there, and in the random forests analysis, where they were entered as numerical covariates to allow for more sensitive interactions along their continuous values. The persistence predictor was seemingly irrelevant in its individual glmer analysis, but it rose to importance when an interaction term was included that tapped the identicality of the prime and target’s auxiliary form. The persistence predictor was also deemed important in the random forests analysis.

15.1.2 Explanations

While we have gained substantial insight into the predictors that effect PAC serial variability in this study, the underlying causes of it are nonetheless still very uncertain. Recall that one should not assume that association is the same as causation.

A number of speculative explanations have been mentioned throughout this thesis, centering on some fairly obvious ones — feature-checking theory, complexity theory, and optimization theory. For feature-checking theory based on a hierarchically-structured syntax, constituents in and around the PAC may move around to satisfy syntactic “checking” requirements (e.g. Chomsky 1995), and so result in the serialization differences we find. However, if this is true, syntactic feature checking is not really a requirement, but more of a “soft” constraint; in addition, it should be noted that the generative enterprise does not find the notion of probabilities in language attractive (see Section 3.2).

As for complexity theory, we assume that less complex items (that is, those that are easier to access) are more likely to be put earlier (see MacDonald 2013). If the participle has a feature that is less complex, say [+SG], it can be produced earlier, whereas if it has a more complex feature, such as [+PL], it takes more time to access, and is thus placed later in the clause. The auxiliary “buys a bit more time” to access/compile and produce the participle.

The optimization theory seems to help account for some of the phonology predictors, such as harmonic rhythmical structure, metrical well-formedness, and the avoidance of segmental clashes. It may also be applicable to information structure. When the preverbal constituent is highly focal (such as being indefinite, contrastive, or topic-worthy), the auxiliary is brought into service to prevent the former and the participle — which is also presumably informationally heavy — from being adjacent to each other.

It may be that all of these are of some explanatory value in tandem. It must be noted that, because the focus of this thesis was on statistical modelling, further explorations are therefore required, and what has been mentioned here and above are no more than tentative speculations.

15.1.3 Methodological conclusions

Along the methodological plane, it has been found that by using a suite of statistical toolkits — such as GLMS, GLMERs, GAMs, conditional inference trees, and random forests — we can enhance our understanding of the complex structure of language variation in the past. GLMERs are effective at showing
how predictors interact with multilevel structure, such as verb lemma effects, which cannot readily be incorporated into random forests. Random forests build many conditional inference trees on bootstrapped samples with a prespecified number of predictor variables available for splitting at each node. Consequently, they allow weaker predictors to shine whose effect would otherwise have been swamped by other predictors higher up on the pecking order. Further, by comparing the variable importance scores of the latter with $AIC$ deviances, it is possible to come to an understanding of which variables are important as main effects and which are only important as moderators. These moderator effects can then be inspected more closely using conditional inference trees. In this respect, while the quote below from Devine & Stephens (2006: 179) (cf. page 31 in this thesis) is an accurate description of the complexity of the dataset, advanced statistical modelling techniques are able to pull out what is of influence:

The location of the auxiliary is a complex issue for which a vast amount of data is available. It is subject to a number of quite subtle semantic and syntactic conditions, which can sometimes be conflicting and pitted against each other in an optimality calculus according to the number and importance of those favouring postverbal position versus those favouring some preverbal position.

Finally, it was shown in Section 14.3.1 that random forests, in terms of their predictive performance, are on the one hand comparable to a similar technique, such as bagging, which has all predictors available for evaluation at each node, but on the other hand superior to a single conditional inference tree.

We have only examined a slice of the statistical toolkits available here. One might wish to explore the dataset with other classifiers, such as support vector machines (Steinwart & Christmann, 2008). We have barely touched on another statistical subject in this study, namely that of factor level reduction. For example, the predictor of $\text{ClauseType}$ comprised many levels. Might it not be sensible to reduce them with respect to how they affect the response? Advanced ways of examining this are suggested in Gertheiss & Tutz (2010); Tutz (2012). Exploration of these multivariate quantitative methods will allow the modeller to understand the data in even greater depth than has been examined here.

15.1.4 Operationalization

Another contribution of this thesis has been to emphasise the collection and operationalization of variables, both response and predictor alike. It has been noted in several places that theoretical definitions need not match up with operational definitions (such as the information structural notion of “focus”). Operational definitions are ways to more easily tap a variable that is otherwise hard to pin down.

More computationally rich ways of operationalizing the variables examined in this thesis are left for further study. For instance, information structure might also be effectively measured by Shannon information content (Shannon, 1948) and via forward and backward bigram probabilities to map the informational contour of a sequence. Topicality might be measured via the TF-IDF (“term frequency-inverse document frequency”) measure (see Pan & McKeown 1999 and Pan & Hirschberg, 2000 for linguistic applications of these two measures). Such operationalizations were not exploited here because the resources available to the computational Latin linguist are somewhat sparse at present.

15.2 Practical applications

There are a number of practical applications to probabilistic and multivariate models of grammar that derive from a historical corpus of a dead language. Below two potential applications are suggested.

15.2.1 Textual criticism

It was noted in Section 6.3 that in the course of the transmission of a text different textual readings can arise in different manuscripts that ultimately derive from the same $\text{prima manus}$, the first hand copy.
which has usually not survived. This is particularly the case with grammatical variation and word order. Devine & Stephens (2006: 32) refer to a study by Rösch (1914), which amasses a great deal of evidence on word order variability in Cicero due to differences in textual transmission, including serialization differences in the \textit{pacc} alternation (1914: d). I submit that probabilistic models have an application in this area of philology, in that they can be used to assist textual critics in deciding on the most likely reading.

15.2.2 Literary stylistics

Probabilistic models are not without literary and stylistic uses, either. For instance, one practical application might be to examine questions of authorship, which has hitherto only been performed via impressionistic stylistic comparisons. The authorships of the \textit{Bella} considered in this study (\textit{Bellum Africum}, \textit{Bellum Alexandrinum}, \textit{Bellum Hispaniense}) are disputed. In antiquity, it was thought that Hirtius — one of the authors considered in this study — was responsible for writing the three \textit{Bella}. However, the \textit{Oxford Classical Dictionary} (Hornblower et al., 2012: s.v.) observes that for the \textit{Bellum Alexandrinum} “stylistic comparison with \textit{Bellum Gallicum} makes his authorship quite possible”, but for the \textit{Bellum Africum} “[t]he attribution to A. Hirtius . . . is impossible”. However, all four works \textit{Bellum Africum}, \textit{Bellum Alexandrinum}, \textit{Bellum Hispaniense}, and \textit{Bellum Gallicum} do not differ in their overall rates of \textit{factus est} and \textit{est factus} in a statistically significant way (as evidenced by a conditional inference tree with \textit{Text}) as the predictor. Similarly, nor do they exhibit any significant interactions with any of the predictors. These statistical facts indicate to all intents and purposes that they are very similar on stylistic grounds with regard to the \textit{pacc} alternation, at least. Assessing via statistical models the four texts’ similarities and differences with regard to a whole range of variable grammatical constructions will allow a more quantitatively detailed picture to emerge concerning the authorship of these works being attributed to Hirtius.

15.3 Final remarks

This thesis has only scratched the surface of the probabilistic and multivariate nature of Latin grammar. For one thing, we have only considered one aspect of a constructional family. Further research will be needed to uncover aspects of auxiliary omission, as in example (4b) on page 28, and discontinuity, as in the examples in (8) on page 30. For another, there are many more grammatical variation phenomena in Latin to attack from the perspective advanced here. Be that as it may, I believe the present thesis has afforded countless insights on language variation phenomena in Latin for probing in further study, and I do not believe that they would have shown themselves without the application of advanced statistical modelling.
Bibliography


276


BIBLIOGRAPHY


Appendix A

Textual Issues

In the following paragraphs I go through some (selected) textual issues, grouped thematically. In a number of cases the editors make different decisions with respect to word order choices concerning the variant (e.g. *factus est* vs. *est factus*, and sometimes *est...factus*). At *B. Afr.* 5, *dum in ea re Caesar esset occupatus*, Klotz (1927) notes two other readings besides this one found in W (codex Laurentianus) and ρ (agreement of codex Vaticanus and codex Riccardianus): *caesar est occupatus*, with a tense and mood change, is found in π and in L and *est caesar occupatus*, with a further difference in word order (engendering an *est...factus* variant), is found in S. Bouvet & Richard (1997) go for *esset occupatus* with Klotz, whereas Du Pontet (1900) opts for the reading in π and L. At *B. Alex.* 53. 5 *erat ibi confecta* where Klotz opts for an *est...factus* variant following T (codex Parisinus), the other editors opt for the *ibi erat confecta* reading, inducing an *est factus* variant. Moving on to Caesar, at *Gal.* 1. 12. 4, Hering (1997) prints *nam omnis ciuitas Helvetia in quattuor pagos divisa est* ‘For the whole state of Helvetia is divided into four cantons’. This *factus est* variant *divisa est* is also adopted by Du Pontet (1900) and Klotz (1952), whereas Seel (1961) opts for *est divisa*, which is present in manuscripts belonging to the π branch in the transmission. Another example is *esse redactas* in *certior factus est omnes eas civitates in dicionem potestatem populi Romani esse redactas* ‘he was informed that all these states had been brought back under the control and power of the Roman people’ (Caes. *Gal.* 2. 34. 1), which is adopted by Hering (1997) and Du Pontet (1900), a variant robustly present in the α branch of manuscripts. An alternative *redactas esse* variant is attested in manuscripts belonging to the β branch, and is the one adopted by Klotz (1952) and Seel (1961). Seel’s (1961) reasoning for preferring *redactas esse* here is that *esse redactas* produces a heroic clausula which, he asserts, never occurs in Caesar. With respect to Cicero, Fedeli (1986) notes, on *Phil.* 2. 117 *comparandus es*, that D has the variant *es comparandus*. The editors of this edition opt for this *est factus* variant, while Ramsey et al. (2009) go for *comparandus es* like Fedeli himself. At Nep. *Alc.* 4. 3, we find *et in trierem, quae ad eum deportandum erat missa, ascendit* in Marshall (1991) and thus in LLT-A. This is to be contrasted with *et in trierem, quae ad eum erat deportandum missa, ascendit*, which Malcovati (1944) opts for. The reading *erat...missa* occurs in a 1608 Frankfurt edition, and was erroneously cited from a lost codex from the 12th century; by contrast, the former reading *erat missa* has the support of other manuscripts.

Other textual concerns involve not word order as such, but rather ellipsis of the auxiliary. For example, in *milites...est cohortatus* ‘he harangued his soldiers’ (Caes. *Gal.* 2. 21. 2) the auxiliary is present in β, and is the reading opted for by Hering (1997). In contrast, α omits the auxiliary, which is the reading adopted by Du Pontet (1900); Klotz (1952); Seel (1961). In another example, Hering (1997) and Du Pontet (1900) overtly realize the auxiliary in *atque inde in Galliam translatae esse existimatur* ‘and it is believed that it was transferred from there to Gaul’ (Caes. *Gal.* 6. 13. 11), whereas Klotz (1952) and Seel (1961) omit it.

Finally, sometimes the lemma of the participle is at issue. Thus, for instance, the *pac coniecta esset*
in *cum hostium acies a sinistro cornu pulsa atque in fugam coniecta esset* ‘when the enemy’s battle line had been routed by the left wing and sent into flight’ (Caes. *Gal.* 1. 52. 6) is the choice of Hering (1997); Du Pontet (1900); Klotz (1952); Seel (1961). Du Pontet (1900), by contrast, opts for *conuersa esset*, a reading in ρ.