

**Quantitative perspectives on variation in Mennonite Plautdietsch**

by

Christopher Douglas Cox

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics  
University of Alberta

© Christopher Douglas Cox, 2015

## **Abstract**

In many endangered language communities, aspects of synchronic linguistic variation (e.g., the extent and formal characteristics of variation, the relationship of identified variables to one another, and the geographical and social distribution of such differences among and across local speaker groups) represent significant gaps in the documentation of local linguistic practices. While such information critically informs current typologies of sociolinguistic variation, its development presents methodological challenges for linguistic research, as limited prior documentation often renders empirically adequate profiles of variation difficult to establish. This is the case in the Saskatchewan Valley, an area in western Canada that served as an important crossroads in the migration of diasporic Russian Mennonite groups throughout the twentieth century. Despite the historical significance and almost unparalleled internal diversity of these communities in Russian Mennonite history, no prior linguistic research has been conducted in this region. This leaves the linguistic consequences of the complex patterns of inter-group contact and separation evinced in these communities poorly understood—a situation with consequences as much for linguistic and historical-cultural research as for community-based language initiatives, where such differences between groups of speakers often also call for attention.

This study addresses the challenges that synchronic variation poses in such contexts through the documentation and analysis of the forms of Plautdietsch (ISO 639-3: pdt) spoken in the Saskatchewan Valley and surrounding areas. Given the absence of previous documentation in this area and the need for such resources in both academic research and community-based language initiatives, this study begins with the development of a Plautdietsch-language primer

(*Fibel*) in partnership with Mennonite and non-Mennonite communities in the region. Through the contributions of several dozen first-language speakers of Plautdietsch, the resulting *Fibel* Corpus serves both as a resource for community language programs and as a standardized survey instrument for assessing the extent and distribution of local linguistic variation in this area. Quantitative, multivariate methods from dialectometry are subsequently applied to these records, identifying and systematically profiling recurrent patterns of variant selection that emerge among groups of speakers. These patterns are then related to the sociodemographic characteristics of the individuals and communities represented, providing a clearer sense of the social and historical embedding of observed variation. The results of this analysis provide not only insights into patterns of linguistic variation of relevance to Mennonite historiography and current sociolinguistic theory, but further suggest the general viability of such community-partnered, documentary, and quantitative approaches to linguistic analysis in contexts of language endangerment and underdocumentation.

## **Preface**

This dissertation is an original work by Christopher Cox. The research project, of which this dissertation is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name “Onse Sproak (“Our Language”): Community-partnered documentation of Mennonite Plautdietsch”, No. 17878, October 5, 2010 (renewed September 16, 2011 and September 27, 2012). No part of this dissertation has been previously published.

## Acknowledgements

Heartfelt acknowledgment is due to each of the individuals from the Plautdietsch-speaking communities in Saskatchewan who contributed to this study. In alphabetical order, this includes Alina (Berg) Balzer, Hella (Neudorf) Banman, Elmer Boehr, Nettie (Harms) Boehr, Kathy (Wiebe) Boldt, Abe Braun, Bill Braun, Dick Braun, Kathy (Guenther) Braun, Margaret (Dyck) Braun, Mary Ann (Friesen) Braun, Abe Buhler, Eva (Schellenberg) Buhler, Jake Buhler, Ruth (Neudorf) Buhler, Wilf Buhler, Leonard Doell, Tina (Janzen) Doell, †Tina (Pauls) Doell, Jack Driedger, Helen Dueck, †Mary (Germain) Freistadt, †Menno Friesen, Tena (Wiebe) Friesen, Anita (Martens) Froese, Elsie Froese, Mary (Driediger) Froese, Julie (Freistadt) Froess, †Frank Funk, Mary (Gerbrandt) Funk, †Jacob G. Guenter, Cornie Guenther, Sarah (Guenther) Guenther, Henry Harms, †Joan (Miller) Harms, Ed Hildebrandt, Mary (Pauls) Hildebrandt, Margaret (Banman) Janzen, †Beth Kobelsky, Bill Kruger, Helen (Zacharias) Kruger, Abe Loewen, †Marie (Siemens) Loewen, Sarah (Siemens) Loewen, Elmer Neufeld, Erna (Suderman) Neufeld, Anna (Guenther) Peters, Bill Peters, Velma (Epp) Regier, Emma (Germain) Schidlowsky, Art Zacharias, Edna (Pauls) Zacharias, and Evelyn (Freistadt) Zerff. Their generosity with time and hospitality extended far beyond what I could ever have hoped, turning even the most ordinary of visits into engaging conversations, often shared around the kitchen table with *Summaborscht* and freshly-baked *Tweeback* or chokecherry pie. Their contributions are at the heart of this dissertation, and reflect as much their tremendous linguistic knowledge as their kindness in sharing it so patiently. As much as I hope that the breadth and depth of their experience is evident in this study, responsibility for any areas that fall short of that goal is ultimately mine alone.

No less substantial and meaningful support has also come from colleagues and advisors in academic communities, both at the University of Alberta and elsewhere. First and foremost, thanks are due to my supervisory committee—Profs. R. Harald Baayen, John Newman, and Sally Rice—for their guidance in all stages of this work. Their generosity in mentorship and deep engagement with this research has left a significant mark, both on this study and on the broader perspectives on linguistic research reflected in it. As my supervisor, Prof. John Newman deserves particular recognition for his guidance as this study took its course, even as it extended through periods of leave and into retirement that should, by all rights, have come with some

reprieve from supervisory responsibilities. The contributions of the external members of the examination committee, Profs. Tom Priestly and William Keel, are also gratefully acknowledged.

Many others have contributed to this work, and deserve acknowledgement here, as well. In Saskatchewan, the Mennonite Historical Society of Saskatchewan opened their archives for research on several occasions, providing information that informed the direction this work has taken. Leonard Doell, Henry Kloppenburg, Q.C., Edward Novecosky, and the late Reuben Epp offered their assistance in making connections with the Plautdietsch-speaking community near Humboldt, Saskatchewan, while Chelsea Cox and Adrienne Findlay helped prepare several recordings for transcription. In Alberta, the Department of Linguistics at the University of Alberta has been a source of inspiration and encouragement since the earliest stages of this project. It is difficult to imagine a more suitable academic home for this work, or to wish for friends and colleagues more creative, supportive, and engaging. I am grateful to Conor Snoek for his introduction to Gabmap and for discussions that led this research in new directions; and to Dr. Martijn Wieling and colleagues at the Eberhart-Karls Universität Tübingen, whose expertise similarly enriched this work during a research visit in September 2012. I am also grateful for the support of friends and colleagues from the Tsuu T'ina Nation, and in particular Elders Bruce Starlight and Doris Roan, who were understanding when this work sometimes meant time away from our regular meetings and ongoing collaborations. In the Yukon, Janet McDonald and colleagues at Yukon Education provided both professional support and encouragement in the final phases of this research. Finally, the Trudeau Foundation community has challenged and broadened my conceptions of linguistic research and its potential for social engagement, often through open-ended (and improbably peripatetic) conversations that have stretched to almost every corner of Canada. I am especially grateful to Dr. Rita Shelton-Deverell for her mentorship, which has been a critical piece of this education. Much of this research has been made possible through the financial assistance of the Trudeau Foundation, the University of Alberta, the Social Sciences and Humanities Research Council of Canada, and the Killam Trusts. I gratefully acknowledge all such support here.

Ultimately, my participation in all of these communities has rested on the constant love and support that friends and family have provided throughout this work. These people, in particular, have offered encouragement when difficult decisions and unforgiving deadlines

loomed, contributed perspective and advised balance when research seemed to call for narrow and exclusive focus, or simply come by with a cup of tea and a conversation at just the right moment. For her patience, encouragement, and faith, and for reflections of these traits in countless kind and beautiful ways over the course of this work, my wife, Olivia Sammons, deserves special recognition here. To my friends, my relatives, my parents, and especially to Livi, for their unfailing love and support—*vālmol Dankscheen*.

## Table of Contents

1. Introduction.....	1
1.1. Mennonites and the Saskatchewan Valley.....	1
1.2. Methodological considerations.....	4
1.2.1. Quantitative sociolinguistics.....	4
1.2.2. Dialectology.....	9
1.2.3. Dialectometry.....	14
1.2.4. Documentary linguistics.....	16
1.2.5. Corpus linguistics.....	20
1.2.6. Summary.....	22
2. Mennonite history and the Saskatchewan Valley.....	25
2.1. Introduction.....	25
2.2. Mennonite origins and migration, 1525–1787.....	25
2.3. Mennonites in Ukraine, 1787–1873.....	29
2.4. Mennonites in North America, 1874–1890.....	31
2.5. Mennonites in the Saskatchewan Valley, 1890–1921.....	32
2.6. Mennonites in the Saskatchewan Valley, 1922–present.....	38
2.7. Summary.....	46
3. Variation in Mennonite Plautdietsch.....	49
3.1. Introduction.....	49
3.2. Hypotheses on Mennonite Plautdietsch differentiation.....	50
3.2.1. Early Mennonite adaptation to local vernacular norms.....	50
3.2.2. Differences between Chortitza and Molochnaya Colonies.....	52
3.2.3. Differences between denominations.....	54
3.2.4. Differences in emigration period and socioeconomic class.....	56
3.2.5. Differences in dialect geography and points of origin.....	60
3.2.6. Summary.....	62
4. Methodology.....	66
4.1. Introduction.....	66
4.2. Research design.....	67



4.2.1. Linguistic items and linguistic variables.....	71
4.2.1.1. Phonemic inventory.....	74
4.2.1.2. Lexical variables.....	86
4.2.1.3. Phonological variables.....	93
4.2.1.4. Morphological variables.....	95
4.2.1.5. Syntactic variables.....	99
4.2.1.6. Summary.....	104
4.2.2. Eene Plautdietsche Fibel.....	105
4.2.2.1. Motivations.....	105
4.2.2.2. Design and review.....	110
4.2.3. Developing the Fibel Corpus.....	118
4.2.3.1. Contributions and contributors.....	118
4.2.3.2. Tasks and recording.....	125
4.2.3.3. Transcription.....	131
4.2.3.4. Coding.....	134
4.2.3.5. Summary.....	137
5. Analysis.....	139
5.1. Introduction.....	139
5.2. Dialectometric analysis.....	141
5.2.1. Phonological clustering.....	147
5.2.2. Lexical clustering.....	169
5.2.3. Morphological clustering.....	182
5.2.4. Syntactic clustering.....	198
5.2.5. Aggregate clustering.....	211
5.3. Conclusions.....	220
6. Conclusion.....	232
6.1. Methodological aims.....	232
6.2. Variation in Mennonite Plautdietsch revisited.....	238
6.3. Linguistic variation and social structure.....	245
6.4. Conclusion.....	248

References.....	250
Appendix A: Mennonite communities in the Saskatchewan Valley.....	271
Appendix B: Non-Mennonite communities in the Saskatchewan Valley.....	275
Appendix C: Linguistic items.....	277
Appendix D: Contributor information forms.....	282
Appendix E: Transcription conventions.....	286

## List of Tables

Table 1. Population of Saskatchewan Valley communities and numbered rural municipalities (2011 Canadian Census). Estimates the total local population, the number of mother tongue speakers of varieties of German (L1), the number of speakers who use these varieties regularly in the home (HL), and a derived rate of overall language retention (HL / L1).....	44
Table 2. Plautdietsch consonant phonemes.....	75
Table 3. Phonemic analyses of Plautdietsch monophthongs. Varieties are listed by country, period of emigration from Russia or Ukraine, and colony of origin (C = Chortitza Colony, M = Molochnaya Colony, <C/M = Russian-Ukrainian daughter colony of Chortitza/Molochnaya)....	78
Table 4. Phonemic analyses of Plautdietsch diphthongs. Varieties are listed by country, period of emigration from Russia or Ukraine, and colony of origin (C = Chortitza Colony, M = Molochnaya Colony, <C/M = Russian-Ukrainian daughter colony of Chortitza/Molochnaya). Where noted, distinctive pre-velar allophones of /ɔa/ and /oa/ are given after a semicolon in the corresponding column.....	81
Table 5. Front diphthong contrasts in Plautdietsch varieties.....	82
Table 6. Lexical exceptions to front diphthong contrasts in Plautdietsch varieties.....	82
Table 7. Back diphthong contrasts in Plautdietsch varieties.....	83
Table 8. Phonemic categories and allophones of Plautdietsch.....	86
Table 9. Lexical variables and corresponding items in the Fibel Corpus.....	88
Table 10. Lexical-phrasal variables and corresponding items in the Fibel Corpus.....	89
Table 11. Lexical-phonological variables and corresponding items in the Fibel Corpus.....	92
Table 12. Macro-variables and corresponding items in the Fibel Corpus representing recurring	

patterns across multiple lexical-phonological variables.....93

Table 13. Phonological variables and their corresponding items in the Fibel Corpus. Starred items coincide with other variables in the phonemic inventory.....94

Table 14. Morphological and morphosyntactic variables and corresponding items in the Fibel Corpus.....96

Table 15. Macro-variables and corresponding items in the Fibel Corpus representing recurring patterns across multiple morphological variables.....98

Table 16. Design of the Fibel.....117

Table 17. Contributors to the Fibel Corpus. Contributors' gender is indicated in the first letter of each anonymous identifier, and contributor ages are given as of the time of their participation. Starred contributors are only sparsely represented in the corpus. Denominational affiliations for contributors and their parents (in parentheses) use the following conventions: B = Bergthaler, CMC = Chortitzer Mennonite Conference, E = Evangelical, FEBC = Fellowship of Evangelical Bible Churches, MBr = Mennonite Brethren, MC = Mennonite Church Canada, OC = Old Colony, RC = Roman Catholic. Contributors' places of birth outside of the Saskatchewan Valley are presented in italics. Contributors' parents' places of birth are divided by country and region: CAN = Canada, with MB = Manitoba (East or West Reserve), SK = Saskatchewan, and WR = West Reserve, Manitoba; UKR = Ukraine, with Ch. = Chortitza Colony, <Ch. = daughter colony of the Chortitza Colony, Mo. = Molochnaya Colony, Ma. = Mariupol region; and USA = United States, with MN = Minnesota and NE = Nebraska.....122

Table 18. Contents of the Fibel Corpus. Contributors here include all individuals present for the Fibel task, with individuals in parentheses not providing a full translation in the given recording, and starred contributors being sparsely represented in the corpus overall.....129

Table 19. Distance-based cluster determinants, phonological-phonetic variables.....162

Table 20. Distance-based cluster determinants, lexical variables.....	175
Table 21. Distance-based cluster determinants, morphological variables.....	185
Table 22. Distance-based cluster determinants, syntactic variables.....	206
Table 23. Aggregate cluster determinants, light blue (central) cluster.....	215
Table 24. Aggregate cluster determinants, dark green (central) cluster.....	215
Table 25. Aggregate cluster determinants, light green (transitional) cluster.....	218
Table 26. Aggregate cluster determinants, light pink (peripheral) cluster.....	218
Table 27. Aggregate cluster determinants, dark blue (Humboldt Catholic) cluster.....	219
Table 28. Sibling contributors and their clusters under aggregate hierarchical cluster analysis. Distances between sibling pairs are given in the number of intervening binary divisions and speakers.....	224
Table 29. Summary of clustering analyses by speaker. (Group labels: C = Kanadier, U = Russländer/USA Molochnaya, C-U = mixed Kanadier-Russländer, H = Humboldt Catholic. POB labels: C = central Sask. Valley, P = peripheral Sask. Valley, O = outside of Sask. Valley, H = Humboldt area).....	227
Table 30. Mennonite settlements in the Saskatchewan Valley, 1892–1908 (after Guenter et al. 1995). Includes organized and unorganized villages, four-corner hamlets, and larger districts of Mennonite settlement, with Dominion Land Survey and latitude-longitude coordinates from Guenter et al. (1995) and SaskGrid 2010 GIS datasets.....	273
Table 31. Non-Mennonite-established towns in the Saskatchewan Valley, after Guenter et al. (1995) and McLennan (2008). Population estimates taken from the 2006 and 2011 Canadian Census.....	275

Table 32. Variable linguistic items by category (lex: lexical selection involving single lexemes, lex-phon: lexically restricted phonological patterns, lex-phr: lexical selection involving phrases and multi-word expressions, phon: general phonological patterns, morph: morphological patterns (case inflection, paradigm selection)). Items are given with attested variants (in the Sass orthography or in IPA, according to the corpus coding conventions for each item), as well as identifiers and corresponding sentences in the Fibel Corpus on which items are coded.....281

Table 33. Transcription conventions in the Fibel Corpus.....287

## List of Figures

Figure 1. Saskatchewan Mennonite reserves (reproduced from Doell 1987: 8).....	2
Figure 2. Vistula delta region.....	27
Figure 3. Proportion of first-language (L1) speakers of varieties of German in the Saskatchewan Valley (left panel), and rates of retention of German as a language of regular use in the home among these speakers (right panel; 2011 Canadian Census).....	45
Figure 4. Density plot of contributor ages.....	124
Figure 5. KML layer of the Saskatchewan Valley in Google Earth. Inset coloured areas show the historical boundaries of the Hague-Osler Mennonite Reserve, with the original 1895 allotment in yellow, and later additions shown in orange (early 1898), blue (August 1898), and purple (October 1898).....	143
Figure 6. Contributors by place of birth (n = 42).....	146
Figure 7. Distribution of vCanPLVowel, with [e] (kjänne(n)) on left and [ɛ] (kjenne(n)) on right. .....	148
Figure 8. Distribution of vRealizationOaPreVelar, with non-fronted realizations (e.g., [oə], [o·]) on left and fronted realizations (e.g., [øø], [eø]) on right.....	148
Figure 9. Distribution of vUe, with back realizations (i.e., [u]) on left and front realizations (i.e., [y]) on right.....	149
Figure 10. Distribution of vEea (left, [eə(r)] in blue) and vOo (right, [o·] in blue).....	150
Figure 11. Density plot of distances to nearest geographical and linguistic neighbours, phonemic-phonological variables.....	151
Figure 12. Reference-point map for contributor F01.....	153

Figure 13. Reference point map for contributor F15.....	154
Figure 14. Reference point map for contributor M00.....	154
Figure 15. Two-dimensional MDS plot, phonemic-phonological variables ( $r = 0.96$ ).....	155
Figure 16. Hierarchical clustering by phonemic and phonological features (Ward's method, six clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.96$ ; right).....	157
Figure 17. Two-dimensional MDS plot (six clusters, phonemic and phonological variables, $r = 0.95$ ), Mennonite speakers only.....	158
Figure 18. Hierarchical clustering by phonemic and phonological features (Ward's method, four clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.96$ ; right).....	159
Figure 19. Two-dimensional MDS plot (four clusters, phonemic-phonological variables), Mennonite speakers only ( $r = 0.95$ ).....	160
Figure 20. Dendrogram of hierarchical clustering by phonemic and phonological features (Ward's method, four clusters).....	161
Figure 21. Distribution of phonemic-phonological cluster determinants, dark blue cluster. Clockwise from top left: vWasVowel ([eə] in blue, [iə] elsewhere), vOo ([oː] in blue, [əʊ] elsewhere), vHorsesVowel ([eə] in blue, [iə] elsewhere), vEe ([eː] in blue, [əɪ] elsewhere).....	163
Figure 22. Distribution of phonemic-phonological cluster determinants, light blue cluster. Clockwise from top left: vAuEeGave (<au> in blue, <ee> elsewhere), vNj ([ɲ] in blue, [ŋ] elsewhere), vAuj ([aʊ] in blue, [ɔj] elsewhere), vAuEeOoTook (<au> in blue, <ee> and <oo> elsewhere).....	164
Figure 23. Distribution of phonemic-phonological cluster determinants, light green cluster. Clockwise from top left: vOea ([oə] in blue), vRealizationUaPreVelar (back realizations in blue, front realizations elsewhere), vUea ([uə] in blue), vRealizationFinalRen (non-metathesized -re(n))	



in blue, metathesized -er(n) elsewhere)..... 166

Figure 24. Distribution of phonemic-phonological cluster determinants, dark green cluster.

Clockwise from top left: vRealizationUaPreVelar (front realizations in blue, back realizations elsewhere), vRealizationOaPreVelar (front realizations in blue, back realizations elsewhere), vOea ([eq] in blue), vCould2SReduction (reduced ku's(t) in blue, non-reduced kunnst elsewhere)..... 167

Figure 25. Density plot of distances to nearest geographical and linguistic neighbours, lexical variables..... 170

Figure 26. Two-dimensional MDS plot, lexical variables ( $r = 0.91$ )..... 171

Figure 27. Hierarchical clustering by lexical features (Ward's method, four clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.91$ ; right)..... 172

Figure 28. Dendrogram of hierarchical clustering by lexical features (Ward's method, four clusters)..... 174

Figure 29. Distribution of lexical cluster determinants, dark blue cluster. Clockwise from top left: vHouse (Kot in blue, Huus elsewhere), vUncles (Oohms in blue, Onkels elsewhere), vAunts (Mumms in blue, Taantes elsewhere), vLate (spod in blue, lot elsewhere)..... 176

Figure 30. Distribution of lexical cluster determinants, light green cluster. Clockwise from top left: vUntil (bat in blue, bott elsewhere), vEnglish (Englisch in blue, Engelsch elsewhere), vBecause (wiels in blue), vGirls (Määkjes in blue, Me(r)jalles and Me(r)jalle(n) elsewhere).. 178

Figure 31. Distribution of lexical cluster determinants, light blue cluster. Clockwise from top left: vUntil (bott in blue, bat elsewhere), vEnglish (Engelsch in blue, Englisch elsewhere), vGirls (Me(r)jalle(n)/Me(r)jalles in blue, Määkjes elsewhere), vSay (saje(n) in blue, saije(n) elsewhere).  
..... 180

Figure 32. Distribution of lexical cluster determinants, dark green cluster. Clockwise from top

left: vThatDay (dän Dag in blue, däm/dee Dag elsewhere), vMoon (Mon in blue, Mond elsewhere), vSome (waut in blue), vWatermelon (Rebuus in blue).....	181
Figure 33. Hierarchical clustering by morphological features (Ward’s method, four clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.93$ ; right).....	183
Figure 34. Dendrogram of hierarchical clustering by morphological features (Ward’s method, four clusters).....	184
Figure 35. Density plot of distances to nearest geographical and linguistic neighbours, morphological variables.....	185
Figure 36. Distribution of morphological cluster determinants, dark green cluster. Clockwise from top left: vEnNounSg, vEn3PL, vEnPastPart, vEnInf (-e in blue).....	187
Figure 37. Distribution of morphological cluster determinants, dark blue cluster. Clockwise from top left: vTheAccMSg (dee in blue, däm/dän elsewhere), vEnWithout (-e in blue), vEnOther (-e in blue), vEndingAdjAccMSg (-a in blue, -en elsewhere).....	189
Figure 38. Distribution of morphological cluster determinants, light blue cluster. Clockwise from top left: vEn3PL, vEnVerbal, vEnNounSg, vEnPastPart (-en in blue).....	191
Figure 39. Distribution of morphological cluster determinants, light green cluster. Clockwise from top left: vEnWithout, vEnOther, vEnNominal, vEnNounSg (-en in blue).....	192
Figure 40. Density plot of distances to nearest geographical and linguistic neighbours, syntactic variables.....	201
Figure 41. Two-dimensional MDS plot, syntactic variables ( $r = 0.98$ ).....	202
Figure 42. Distribution plots of verb raising (left) and verb projection raising (right) constructions.....	203

Figure 43. Hierarchical clustering by syntactic features (Ward’s method, three clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.98$ ; right).....	204
Figure 44. Dendrogram of hierarchical clustering by syntactic features (Ward’s method, three clusters).....	204
Figure 45. Hierarchical clustering by syntactic features (Ward’s method, two clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.98$ ; right).....	205
Figure 46. Dendrogram of hierarchical clustering by syntactic features (Ward’s method, two clusters).....	206
Figure 47. Distribution plot of verbal complement orders (1-2 in blue).....	210
Figure 48. Density plot of distances to nearest geographical and linguistic neighbours, all categorical variables.....	213
Figure 49. Hierarchical clustering on all categorical features (Ward’s method, five clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.94$ ; right).....	214
Figure 50. Dendrogram of hierarchical clustering on all categorical features (Ward’s method, five clusters).....	215
Figure 51. Hierarchical clustering on all categorical features (Ward’s method, five clusters; left) and a distribution plot of vOea (right, [eq] in blue).....	216
Figure 52. Hierarchical clustering on all categorical features (Ward’s method, five clusters; left) and a distribution plot of vEnInf (right, -e in blue).....	220
Figure 53. Dendrogram of hierarchical clustering by all categorical features (Ward’s method, five clusters).....	223
Figure 54. Mennonite settlements in the Saskatchewan Valley, 1892–1908 (after Guenter et al.	

1995).....274

Figure 55. Non-Mennonite-established towns in the Saskatchewan Valley, after Guenter et al. (1995) and McLennan (2008).....276

Figure 56. Contributor information form, personal and family information.....283

Figure 57. Contributor information form, language proficiency information.....284

Figure 58. Contributor information form, language use information.....285

## **1 Introduction**

### *1.1 Mennonites and the Saskatchewan Valley*

Between 1895 and 1930, several thousand Russian Mennonites—members of Anabaptist Christian denominations with historical ties to the lowlands of northern Europe and the Ukrainian steppes—arrived as settlers on the prairies of central Saskatchewan in western Canada. Through negotiation with the Canadian government, tracts of land had been reserved for Mennonite colonization in the Saskatchewan Valley, a geographical basin formed by the approach of the North and South Saskatchewan Rivers in central Saskatchewan before their confluence east of Prince Albert. Over time, newcomers from diasporic Mennonite settlements in Manitoba, the midwestern United States, eastern Europe, and western Siberia came to converge on the small region of the central Saskatchewan Valley shown in Figure 1, where it is referred to as the Hague and Rosthern Mennonite Reserves.

These Saskatchewan Valley settlements served as an important crossroads for Russian Mennonite migration throughout the early twentieth century, acting both as a meeting place of historically separate groups and as a point of departure for subsequent migrations that would establish significant Mennonite settlements throughout the Americas. Indeed, together with Mennonite communities in southern Manitoba, with which the Saskatchewan Valley shares extensive historical connections, these western Canadian settlements are largely without parallel in Russian Mennonite history. Few areas of present-day Mennonite settlement can claim as diverse or complex a history of sustained inward and outward migration over the course of successive generations, or arguably bear equally vivid witness to the consequences of each such wave of migration for the communities involved. Multiple historical ‘layers’ of migration, interaction, and isolation between Russian Mennonite groups are still perceptible in the current diversity of the Saskatchewan Valley settlements, offering an almost synecdochic reflection of the Russian Mennonite story as a whole.

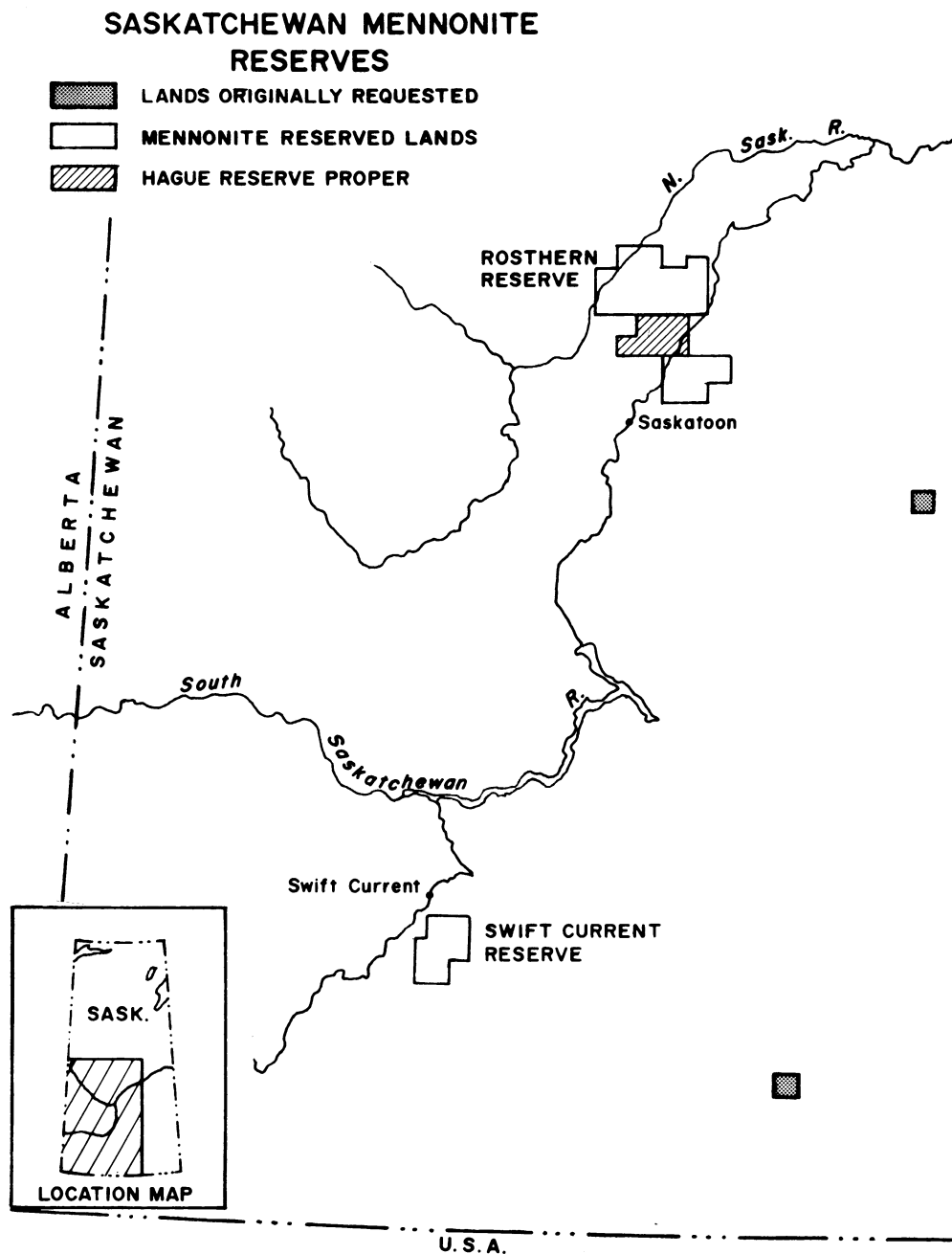


Figure 1. Saskatchewan Mennonite reserves (reproduced from Doell 1987: 8)

The effects of the continual, diasporic migrations of a “wandering and pilgrim people” (Guenter et al. 1995: 446) are no less evident in variation observed in the forms of Mennonite Plautdietsch (Mennonite Low German, *Mennonitisches Niederdeutsch*; ISO 639-3: pdt) maintained in these communities today. The aim of the present study is to investigate such

variation, developing initial documentation of differences in those forms of Plautdietsch represented in central Saskatchewan in partnership with these communities and demonstrating the applicability of the resulting records to contemporary linguistic analysis through a description of the synchronic distribution of such variation across speakers and communities in the region.

At the heart of this study, then, are problems that such patterns of contact and separation between groups in the Saskatchewan Valley present for linguistic investigation, especially given the dearth of documentation concerning the linguistic practices of these communities. Although first-hand observation and informal reports suggest considerable linguistic variation to be present in the region, no prior linguistic research has been undertaken in this area. In the case of the Saskatchewan Valley, such problems lead to the formulation of two general questions that guide this research: how might one effectively identify, systematically document, and adequately describe the linguistic profiles of an unknown number of varieties that may be present in a speech community? Moreover, how might these analytical goals be pursued in consonance with the perspectives and priorities of the communities involved, such that these inform the conduct of research and encourage outcomes that are of benefit to all parties?

Although this study focuses on these Saskatchewan communities in particular in addressing these questions, similar problems are increasingly common elsewhere in linguistics, as well. With the rise in academic linguistic interest over the past twenty years in the description and documentation of minority and Indigenous languages, the range of problems that emerge when seeking to develop adequate records of linguistic conventions in such communities, particularly in the absence of prior information on either these practices or the patterns of variation manifested in them, have assumed greater prominence in the discipline. As the following sections consider in greater detail, linguistic variation bears importantly on such documentary and descriptive undertakings, and represents one area where the results of this study may be of particular relevance to current linguistic practice.

The remainder of this study is organized as follows. To establish the academic context and overall direction for the research to come, Section 1.2 considers the perspectives of several linguistic subdisciplines with an interest in synchronic linguistic variation. Chapter 2 provides further information on the history of the Mennonite communities in the Saskatchewan Valley, proceeding from their origins in the Radical Reformation of the sixteenth century up to the

present day. This historical-contextual information is taken up in Chapter 3, which presents an overview of previous research on linguistic variation in the Russian Mennonite diaspora, summarizing dominant hypotheses concerning the range of factors relevant to such variation's present-day distribution. Bearing these reports and the aforementioned disciplinary perspectives on variation in mind, Chapters 4 and 5 lay out the design and implementation of the present study in the Saskatchewan Valley, describing in detail the development of a specialized corpus of Mennonite Plautdietsch and its use in quantitative analysis of synchronic patterns of variation. Finally, these results are discussed in Chapter 6, comparing them against the initial goals of this study, considering their relevance to current models of linguistic variation in Russian Mennonite communities and elsewhere, and identifying potential directions for further research.

## *1.2 Methodological considerations*

The descriptive and documentary problems raised by linguistic variation in the Saskatchewan Valley fall within the scope of several distinct linguistic subdisciplines, each offering its own perspective on how the analysis of variation is generally to be approached. This section discusses several research areas that are closely aligned with the above problems, including forms of variationist sociolinguistics (§1.2.1), dialectology and dialectometry (§1.2.2, §1.2.3), documentary linguistics (§1.2.4), and corpus linguistics (§1.2.5). As much interrelation as these areas may have, there are also points in which their theoretical perspectives differ and where the recommendations each would make concerning research practice diverge. Bearing such differences in mind, this section seeks to draw on the perspectives of these subdisciplines selectively to inform the choice of methods for investigating linguistic variation in the Saskatchewan Valley.

### *1.2.1 Quantitative sociolinguistics*

Since an account of linguistic variation and its possible correlation with historical and contemporary demographic features of Russian Mennonite communities lies at the centre of this research, the relevance of variationist approaches to sociolinguistics seems readily apparent. The concern of such research for the systematic nature of linguistic variation and its investigation of social and attitudinal correlates of observed linguistic behaviour are clearly relevant to the



present exploration of linguistic variation in the Saskatchewan Valley. While this discussion focuses primarily on the quantitative, Labovian variationist tradition in sociolinguistics, the extensive literature in other areas of sociolinguistics may also have contributions to make to research such as this. In particular, perceptual dialectology (e.g., Preston 1989, 1999) and both quantitative and ethnographic-qualitative research into linguistic ideologies and attitudes towards language may shed light on attitudinal boundaries between constituent speaker groups and the social markedness of certain linguistic features that may be less evident from observation of linguistic variants alone. As the concentration of this study is not on attitudes towards variation *per se*, but rather on the formal characteristics of linguistic variation and its correlation with sociohistorical features of the Saskatchewan Valley, the investigation of local language attitudes is left as a topic for further investigation, one which might complement the present focus with additional insight into the social embedding of the linguistic divisions identified here.

This research shares affinities with variationist sociolinguistics in several respects. Beyond the common concern with synchronic patterns of variation in the Saskatchewan Valley, the central role played by corpora of observed language use in variationist sociolinguistic work is much in line with the descriptive and documentary goals of the present study, as is the practice of modelling linguistic variation on the basis of such information. Likewise, much can be gained from the concept of the *linguistic variable*—a construction with multiple possible instantiations whose occurrence may be conditioned by both linguistic and non-linguistic factors (Labov 1972). As we will see in Section 4.2.1, this notion of a single variable subsuming several possible variants is important and presents one particularly useful means of modelling linguistic variation here.

Although the focus of this study on observed linguistic variation and its social correlates places it in close company with many of the traditional emphases of quantitative variationist sociolinguistic research (cf. Tagliamonte 2006), there are also clear points of divergence between the goals of this research and common sociolinguistic practice. It is not uncommon for variationist studies to focus on a relatively small number of linguistic variables and social factors. Typically, these features of interest are selected in advance and compared against a limited set of sociodemographic categories hypothesized to be of explanatory value (cf. Allen et al. 2007, Meyerhoff & Nagy 2008). This concentration on a restricted range of phenomena and

conditioning factors has the benefit of encouraging a thorough treatment of the variation under consideration. At the same time, however, the relationship between this consequently well-studied variation and other instances of variation in the speech community may remain unclear, leaving questions as to the degree to which linguistically significant social divisions observed in the few investigated forms can confidently be treated as representative of larger trends in the speech community, and not merely idiosyncrasies of the variables under consideration.

Reflecting on the range of variables to include in her study of linguistic variation in East Sutherland Gaelic, Dorian (2010: 113) comments relatedly that

[w]ithin the chosen sphere of variation, I have been concerned not to narrow the field of variables unduly. In particular, I have wanted to avoid selecting just two or three variables and following only that tiny selection, because it has seemed to me [...] that sociolinguistic studies have sometimes selectively treated too few variables of too few sorts, and those few in too little relation to each other.

While this “micro-sociolinguistic, quantitative paradigm” (Meyerhoff & Nagy 2008: 14) has been dominant in much quantitative variationist sociolinguistic research, notable exceptions exist. For one, the Tyneside Linguistic Survey (TLS), a sociolinguistic research programme of the late 1960s and early 1970s, sought to determine “the ‘ecology’ of urban varieties of English (that is, what kinds of variation exist), using a radical and rigorous statistical methodology that had evolved in opposition to the already predominant Labovian paradigm” (Allen et al. 2007: 17). In contrast with the largely top-down paradigm prevalent in contemporary Labovian variationist sociolinguistics, the TLS set as its aim a bottom-up exploration of both linguistic and non-linguistic features, their respective clustering, and possible correlations between them:

Rather than pre-selecting ‘salient’, linguistic variables and correlating these with a narrow range of external indices, such as social class, the TLS grouped speakers and analysed their similarity to one another by comparing their data sets across a multitude of variables simultaneously. Each informant would thus be assigned a unique position in linguistic ‘space’, and differences between speakers would be evident in the manner in which these clustered relative to one another. ‘Linguistic’ clusters (grammatical, phonological and prosodic variants) could then be mapped onto ‘social’ clusters, likewise arrived at by multivariate analyses of the subjects’ scores on a wide range of social and lifestyle factors from ‘educational level’ to ‘commitment to taste in décor’.

As Allen et al. (2007: 17) note, while this distinctive theoretical and methodological approach to

exploring socially situated variation “aroused a certain amount of interest” in the scholarly community, it was perceived at the time of its development to be overly complex, devoting excessive attention to detail in methodology and theory that detracted from the overall strength of its results. Ultimately, the Survey’s research programme was left unfinished, and was largely forgotten until mid-1990s, when the development of the Newcastle Electronic Corpus of Tyneside English brought renewed attention to this important, if somewhat neglected, precedent to contemporary multivariate methodologies in quantitative sociolinguistics.

As well, the concentration of this research on an underdocumented minority language, one maintained in a multilingual setting where significant language shift is in progress, stands in contrast to the predominance of variationist sociolinguistic studies concerned with largely monolingual groups representing major western European languages.<sup>1</sup> While certainly not without important variationist precedents in research on Mennonite Plautdietsch (e.g., Kaufmann 1997, 2003b; Steffen 2006a, 2006b), this focus on an endangered non-Anglo-Romance language places this study decidedly in the minority in current sociolinguistic research, and at times affects the degree to which recommendations made in that literature can be applied directly to the present study. This is the case with common sociolinguistic practices relating to stratified sampling methods, for instance, in which the speech community is divided *a priori* into demographic subsets for which appropriate linguistic representatives are sought (Sankoff 2005). Such methods may face difficulties not only in determining the demographic properties relevant to analysis without prior sociolinguistic research in the community, but also in the demographic skew that often accompanies advanced language shift. Leaving aside issues of potential social and economic marginalization that may contribute to language shift (and that have consequences of their own for sampling), prolonged interruptions in intergenerational language transmission

---

1 Nagy (2012: 428–429) observes that almost three quarters of recent publications in leading sociolinguistic journals (*Language Variation and Change*, *Journal of Sociolinguistics*) concern either English or Romance languages. This contrasts markedly with other linguistic subdisciplines, where a much less significant Anglo-Romance bias is evident: in phonology, for instance, Nagy reports that over four-fifths of recent publications in the flagship journal *Phonology* were concerned with languages other than English and Romance. See also Meyerhoff & Nagy (2008) on the underrepresentation of multilingual and minority speech communities in mainstream sociolinguistic research, but also Stanford & Preston (2009) for notable instances of quantitative variationist research involving Indigenous minority languages.

often result in speaker populations concentrated almost exclusively in the oldest generation. In such cases, stratification by seemingly basic categories such as age may not be representative of the actual distribution of the target linguistic knowledge and practices in the community, and may actually come at the cost of significantly under-representing these groups of speakers. Circumspection thus appears warranted in considering the application of sociolinguistic sampling procedures that assume the availability of representatives from a full range of possible ages and socioeconomic backgrounds in contexts such as that of the Saskatchewan Valley.

In a similar way, the common use of sociolinguistic interviews as the primary means of observing language use in much variationist research, attempting to address concerns over the possible effect of observation on the resulting data (the Observer's Paradox; Labov 1972), may require adaptation in the case of underdocumented and endangered languages. Tse (2013) discusses several limitations of interview methods for sociophonetic research involving minority, underdocumented languages, noting that standardized interviews must often be supplemented with word lists and reading tasks, as "casual speech does not always produce enough samples of the variable of interest" (131) for later analysis, and that proficiency in the language of investigation is often assumed, which may not always be the case in research involving smaller or endangered language communities. Even when conversational proficiency is not at issue, the problem of pursuing interviews in the language of investigation without first having a sense of the range of varieties present in the speech community or the relationships of social markedness between them is a significant one, since the use of a particular variety by the interviewer may unwittingly influence contributors' responses. Although some studies have attempted to mitigate such effects by pairing interviewers with respondents of similar backgrounds, this, too, is challenged by the lack of prior documentation. The range of varieties present in the community is itself an empirical question, not a fact known at the outset, and attempting to match interviewers to particular groups of respondents merely presupposes many of the linguistic and social divisions that are meant to be an outcome of research.

Issues such as these present motivation for initial, baseline descriptive research to better establish the range of variation present in the community, such that more detailed and potentially more methodologically diverse research can subsequently be undertaken. Even while bearing such areas of divergence in mind, the dedicated attention given in the variationist sociolinguistic

literature to the social correlates of linguistic variation and to attendant methodological issues are relevant to the descriptive aims of this study, and present models of research practice that will be drawn on further in Chapter 4.

### *1.2.2 Dialectology*

Another subfield of linguistics relevant to the present research is dialectology, a branch of sociolinguistics concerned with linguistic variation in its areal dimension. Although a range of topics are subsumed under linguistic variation and areality, as the European Dialect Syntax project observes, dialectological research has historically concentrated on

issues that pertain to the description of [...] variation and that are standardly part of dialectal research: (i) the geographical distribution of linguistic variables, (ii) correlations between distinct linguistic variables and (iii) the connection between linguistic variation and diachronic change.<sup>2</sup>

This emphasis on the interrelationship of geography, linguistic variation, and processes of language change has a particularly long history in dialectological research, and contrasts in this respect with much research in the Labovian sociolinguistic tradition, where investigations of linguistic variation and physical space have been on the whole less prominent than those involving other social factors (Britain 2010).

Although certainly not the only research tradition in dialectology—both perceptual dialectology and dialectometry could also be seen as distinct streams within this field—the specific questions raised by these ‘traditional’ foci of dialectological work are potentially of interest in the context of the Saskatchewan Valley. Dialectological considerations draw particular attention to possible correlations in these communities between physical geography and linguistic variation, between subsets of linguistic variables that pattern similarly, and to the evidence available for processes of language change at work—all questions of descriptive interest and potential explanatory value in understanding the dynamics of linguistic variation in these communities. Unlike the long-standing dialect areas of continental Europe that have been the focus of much dialectological research, however, it is unclear to what extent common dialectological assumptions of meaningful correlation between linguistic variation and features

---

<sup>2</sup> [http://www.dialectsyntax.org/wiki/Chapter\\_1:\\_Introduction#Empirical\\_interests](http://www.dialectsyntax.org/wiki/Chapter_1:_Introduction#Empirical_interests) (retrieved January 25, 2014)

of physical geography might be reflected in the present-day distribution of variation across the Saskatchewan Valley. The Mennonite presence in central Saskatchewan is relatively recent and, as the following chapter describes, significant societal changes over the past half-century have favoured both greater mobility and increased demographic heterogeneity in these communities. These factors call into question the degree to which physical geography may be correlated with the present-day distribution of linguistic features across communities in the region.

Relatively little research has been dedicated to the spatial features of linguistic variation in Mennonite Plautdietsch speech communities. Although several researchers (e.g., Mitzka 1930, among others) have shown an interest in Russian Mennonite varieties of Plautdietsch for the comparative evidence they may provide for historical dialect geography, there has been less engagement in non-comparative studies in other broadly dialectological topics. Outside of the work of Steffen (2006a, 2006b) in Belize, which explicitly considers ‘diatopic’ variation (here, across different Russian Mennonite settlements in that country) among a host of other possible dimensions of variation, and Nyman (1997), who represents variation in western Siberian Mennonite Plautdietsch according to its geographical distribution across the villages of the Orenburg settlement, relatively little research has taken up common dialectological questions of possible correlations between variation, areality, and diachrony.<sup>3</sup>

Studies such as these also demonstrate connections between dialectological interests and research on linguistic enclaves (*Sprachinseln*). Linguistic enclaves refer to geographically bounded settlements of linguistically distinct minorities surrounded by dominant linguistic majorities (cf. Mattheier 1994, Hartman Keiser 2009: 3–4, Putnam 2011). Many studies of such enclaves concentrate on processes of linguistic change and convergence, whether internally motivated or prompted by contact with other speech communities (cf. van Ness 1996, Hartman Keiser 2009), with a particularly long tradition of such scholarship involving German-speaking communities (often running in tandem with dialectological research; cf. Riehl 2010: 335–337). Given a similarly pronounced interest in issues of language contact and change in studies of

---

3 However, it could be argued that the descriptive profiles of variation offered by Dyck (1964) and others present implicit correlations between geographically rooted speech communities and local patterns of linguistic variation, although such studies more often portray communities’ linguistic practices as relating primarily to denomination or emigration history, rather than to physical space proper, and seldom give systematic attention to areal factors as being potentially explanatory of variation in their own right.

Plautdietsch, it is not surprising to note that diasporic Russian Mennonite communities have often been framed as linguistic enclaves, as well, as with Rohkohl (1993) in his description of the *Russländer* Paraguayan Mennonite settlement of Fernheim and Steffen (2006a, 2006b) on Mennonite settlements in Belize.

Yet, the situation of Mennonite Plautdietsch is decidedly different from the models commonly assumed in European dialectology and German enclave studies in several respects. As noted in Chapter 2, Plautdietsch represents only one of the community-internal varieties in the traditional Russian Mennonite diglossia. While language use in *sinndöagsch* or ‘Sunday-like’ contexts is dominated by heavily codified forms of Mennonite Standard German (*Huagdietsch*), much less social markedness is attached to the use of Plautdietsch in *auldöagsch* or ‘everyday’ contexts (cf. Hedges 1996). As Cox (2013) argues, this affords considerable room for linguistic variation and change: here, normative attention is focused primarily on maintaining symbolic control over use of the standard code in *sinndöagsch* contexts, rather than on monitoring potential innovations in the *auldöagsch* sphere. Importantly, under this same arrangement of varieties, Plautdietsch is not taken to be an aberrant form of *Huagdietsch*, but instead stands as an autonomous, ‘roofless’ (*dachlos*; cf. Barbour & Stevenson 1998, Haarmann 2005) set of varieties without a single standard form. In both its typical assignment to socially unmarked and largely community-internal functions and the absence of an acknowledged standard variety, Mennonite Plautdietsch contrasts with many other European languages, whether in linguistic enclaves or elsewhere, whose regionally or socially defined varieties are typically subordinate to a standard language. In the Russian Mennonite case, there is no single variety of Plautdietsch that can be treated as the acknowledged standard. This lack of a single normative point of reference for Plautdietsch language use again places emphasis on understanding extant variation in a somewhat less hierarchical fashion, taking lateral connections between coeval varieties as a major focus of research rather than individual varieties’ departure from or convergence to the model of an assumed standard language.

As elsewhere in sociolinguistics, Chambers & Trudgill (1998: 47) stress the importance of representativeness of linguistic sampling in dialectology, and thus also on the careful selection of speakers from the larger population. On their view, such samples ideally comprise a wider segment of the speech community than only non-mobile, older, rural males (NORMs) who were

the traditional target participants for much earlier dialectological research, given their presumed linguistic conservatism. Chambers & Trudgill (1998: 45–47) point out that such speakers are likely to be atypical of the larger population, and that their overrepresentation in dialect samples may be problematic when it implies the underrepresentation of other segments of the population. As a result, they advocate sampling methods not based on personal contacts, which they find to be an “unreliable method of selecting and obtaining informants” (47), but instead forms of canvassing and other structured, semi-random sampling that aim to achieve even representation of major demographic segments of the population divided by age, gender, class, and similar, predetermined social features. While the motivations behind these recommendations about sampling procedures are entirely reasonable, several tacit assumptions about the typical profile of a speech community in dialectology underlie them, namely that:

- a. community members are readily identifiable in sufficient number across all demographic categories hypothesized to be relevant to variation between dialects;
- b. community members are open to vernacular interactions with individuals outside of their local social networks; and
- c. significant social and economic stratification exists within the community, such that a meaningful analysis of these common sociolinguistic factors is possible.

Considering each of these assumptions in turn, it is no doubt ideal to have representation of all segments of the target speech community in even balance, rather than selecting only a small and potentially atypical demographic sliver. An older cohort of speakers, in particular, may well maintain vernacular usage patterns no longer in currency elsewhere, and may thus present a limited view on what is typical in other parts of the speech community. Yet, this ideal may be difficult to square with the demographic realities of language shift and loss and of minority status in a larger population. As the preceding section has noted, situations of advanced language loss often result in populations of speakers no longer being evenly distributed over the kinds of social categories that Chambers & Trudgill (1998) identify. Rather, the concentration of traditional language use in older age cohorts and limited maintenance in younger generations implies that much linguistic research engaged with such languages in contexts of severe endangerment will inevitably be concerned with the linguistic practices and knowledge of older speakers. While concentration on older populations may be an unwarranted methodological proclivity in other



areas of dialectology, it is often a virtual necessity for such research in speech communities in which profound language shift is underway. These considerations bear directly on the Saskatchewan Valley, as well. In traditional, agrarian Mennonite settlements, many Plautdietsch speakers are likely to be both rural and non-mobile (in the sense of maintaining residency on permanent farmsteads, although certainly marked by larger, community-level migrations and possibly also smaller, family-level relocations involving the acquisition or sale of land). Concentrating on older, rural, non-mobile populations in the context of these communities is arguably not 'atypical' in the sense that Chambers & Trudgill (1998) intend, as comparable populations of mobile, urban, and younger speakers in the same community are not being left unrepresented. Rather, this focus on the primary body of speakers acknowledges the nature of demographic skew in speech communities whose speaker populations have been altered as the result of interrupted intergenerational language transmission.

It is the same minority status and level of endangerment that renders the canvassing methods and other forms of random population sampling advocated by Chambers & Trudgill (1998) problematic here. In many research contexts where speakers of the language(s) of interest form the majority, it may not be difficult to identify sufficient numbers of respondents through these forms of sampling. With many minority languages, however, without the use of surnames or other such information as heuristics, structured, semi-random sampling is unlikely to succeed in reaching the speech communities of interest. Moreover, such sampling methods presume a degree of success in soliciting participation without any prior established relationship of trust (or, indeed, any previous contact). In the contexts of significant linguistic and often social marginalization that commonly accompany language loss, however, as well as in many Russian Mennonite communities generally, traditional vernacular language use is commonly restricted to interactions with intimates, making it at best unusual for individuals not identified through local networks of kinship and interaction to serve as interlocutors.

While offering valuable perspectives on research practices that have proven successful in many situations, such recommendations require adaptation to cases where the relevant contexts of language use are less amenable to observation without consideration of the local networks of relationships within which they exist; where speech communities are in the minority and potentially less identifiable by random sampling; and where the demographic and socioeconomic

make-up of the community may be much less differentiated than in other, more familiar majority language contexts. These considerations are relevant both for what is to be considered ‘representative’ in a given speech community and for the selection of methods that are sensitive to the sociolinguistic situation of the community itself.

Other characteristics of traditional dialectology deserve consideration here, as well. Dialectological research has long focused on lexical and morphophonological variation, often leaving other aspects of linguistic organization to receive less attention. Until the recent advent of linguistic atlases based on sociophonetic methods (e.g., the *Atlas of North American English*; Labov, Ash & Boberg 2006) and large-scale surveys of dialect syntax and prosody (e.g., Barbiers et al. 2005, Barbiers, Cornips & Kunst 2007, Prieto, Cabré & Vanrell 2010), dialectological treatments of subsymbolic (e.g., fine phonetic) variation and features of language use above the level of the word have been uncommon. In part, this reflects the limitations of non-audiovisual representations of language available to earlier studies, as well as the prevalence of controlled linguistic tasks intended to provide comparable data on a predetermined range of phenomena. Much dialectological research has proceeded from the implementation of carefully prepared linguistic questionnaires, whether conducted through fieldwork (e.g., with the renowned efforts of Edmond Edmont for the *Atlas Linguistique de la France*; Edmont & Gilliéron 1902) or through the written responses of untrained respondents (e.g., with the large-scale, sentence-translation-based survey of Georg Wenker for the *Deutscher Sprachatlas*; cf. Lameli 2010: 575–576). More recent dialectological research has also sought to apply a wider range of methods in gathering linguistic data, incorporating acceptability ranking tasks, sentence completion tasks (with visual prompts to constrain the range of possible interpretations), cloze tasks, and other experimental methods into dialectological studies (cf. Barbiers, Cornips & Kunst 2007). The present study may thus be able to draw on the methodological breadth of recent dialectological research, selectively adopting techniques from this area that may be suited to the sociolinguistic situation of the Saskatchewan Valley.

### 1.2.3 Dialectometry

As its name suggests, dialectometry is concerned with the development and application of measures of linguistic distance between dialects. First proposed by Séguy (1971) and Goebel

(1982), these approaches sought to address limitations to the divisions proposed in traditional dialectology, which are often made on the basis of a few hand-selected features, through the application of aggregate measures of distance to entire sets of variables. As an example, where traditional dialect atlases typically based their conclusions about dialect divisions on the mapped distributions of a small number of features argued to be significant, dialectometric analyses of the same data base their analytical divisions not on one or another preselected map, but essentially on the sum-total of all available maps, with each such feature contributing its part to the final, overall classification of varieties.

There are several potential advantages to such an approach to linguistic taxonomy over traditional dialectological methods. As both Goebel (1982: 12) and Nerbonne et al. (2011) point out, the aggregation of all available data points may make patterns of interrelation between features more perceptible, exposing commonalities in their distributions that might otherwise be missed in manual inspection. This emphasis on aggregate measures also offers a safeguard against the criticism raised by Malkiel (1976: 71) against dialectologists' "infatuation with the inexhaustible stock of local idiosyncrasies" that sometimes prevents them from "recogniz[ing] the forest, since they are enthralled with the trees; in fact, by the leaves, the branches, the twigs, the roots and rootlets, the petals, and the pollens." Rather, as Szmercsanyi (2014) suggests,

in dialectology, so-called "single-feature-based studies" [...] are fine when the research question is tree-centered – i.e. when it is the features themselves that are of analytic interest. But single-feature-based studies are inadequate when it comes to characterizing 'forestry', multidimensional objects such as dialects or varieties (or relations between them). [...] So, the aggregate perspective [...] is called for when the analyst's attention is turned to the forest (i.e. the multitude of features that characterize a given dialect), not the trees (i.e. individual features of a dialect). Aggregation mitigates the problem of feature-specific quirks, irrelevant statistical noise, and the problem of inherently subjective feature selection, and thus provides a better description of dialects, and a more robust linguistic signal. This robust linguistic signal also facilitates comparison of different forests (that is, dialects), which is after all a key objective in dialectometry.

The combination of linguistic geography with multivariate statistical techniques found in other disciplines, Goebel (1982: 58) argues, presents one way forward in empirical dialect classification—one in which broad trends in the patterning of geolinguistic features are neither lost in the sheer volume of variation nor obscured by arbitrary classifications made on the basis of single,

hand-picked features. An accompanying emphasis on the use of published algorithms in classification also provides support for independent replication of research, making use of the capacity of contemporary computers to consider more features in more possible configurations than would be possible with hands and eyes alone.

Several aspects of mainstream dialectometry are attractive for exploring the geolinguistic dimensions of local linguistic diversity in the Saskatchewan Valley. The capacity of dialectometric methods for large-scale, simultaneous comparisons of variables representing multiple aspects of linguistic organization, none of which have necessarily been identified in advance as having particular explanatory relevance, might allow significant linguistic characteristics of local varieties to be identified inductively, without prematurely limiting the range of features relevant to their description. At the same time, these methods' orientation towards detailed, computationally accessible information on linguistic variation presents an opportunity to explore intersections with other subdisciplines that commonly provide such multivariate data, such as corpus and documentary linguistics (cf. Szmrecsanyi 2011, as well as §1.2.4 and §1.2.5 below). For the exploration of largely undocumented variation with possible geographical correlates, methods such as these that allow for replicable investigations to be conducted on the basis of permanent documentation are of understandable attraction.<sup>4</sup>

#### *1.2.4 Documentary linguistics*

The preceding discussion has framed the situation in Saskatchewan Valley Mennonite communities as largely a sociolinguistic one and, accordingly, has considered points of connection with variationist sociolinguistics, dialectology, and dialectometry that might inform the methodological direction that this study takes. The same situation can also be viewed from the perspective of documentary linguistics, a subdiscipline concerned with issues raised by the

---

4 As the preceding quotation from Szmrecsanyi (2014) suggests, dialectometric methods also permit comparisons between linguistically related but geographically and historically distinct groups; see Nerbonne & Heeringa (2001) for one such cross-linguistic dialectometric study involving Dutch and western Siberian Plautdietsch. Although beyond the scope of the present study, such methods might thus allow for empirically well-founded comparison across speech communities, giving some sense of their relative similarity or difference in particular aspects of linguistic organization (cf. Spruit 2008, ch. 4) and presenting another means of investigating their historical development.

development and application of permanent, diverse, and accessible collections of linguistic records (Himmelman 1998, 2006, 2008; Woodbury 2003, 2011).<sup>5</sup> Although similarities are evident with the other areas of linguistics surveyed above, perhaps most notably in the common emphasis on the importance of records of observed linguistic behaviour (cf. Flores Farafán & Ramallo 2010, Cox 2011b), documentary linguistics strikes a different profile in several respects:

- Documentary linguistics is broadly concerned with theorizing the development and use of lasting collections of linguistic resources, rather than with specific structural, cognitive, or sociolinguistic questions that might be addressed with these resources. This once-removed stance from the linguistic interests of particular subdisciplines is intentional, and follows from the focus of documentary linguistics on the general relationship of linguistics to the data it produces and relies upon in analysis;
- This concern with the development of linguistic resources and their reuse over a range of contexts also leads to consideration of ethical aspects of linguistic research, particularly as they concern the potential breadth of later applications of documentation. This often involves consideration of the roles of stakeholders in such work, including members of those speech communities whose language(s) are the focus of documentary activities;
- Documentary linguistics commonly makes a distinction between the compilation, annotation, and preservation of unique linguistic records on the one hand (*documentation*) and the application of those records to analytical and practical tasks on the other (*description*). As Himmelman (2012) notes, both the compilation of ‘raw’ linguistic data (e.g., audiovisual recordings, manuscripts) and its later annotation into more accessible primary data (e.g., transcripts, critical editions; cf. Himmelman 2012: 188, 193) are not theoretically neutral tasks, nor devoid of analytical steps that deserve the critical attention that documentary linguistics seeks to give them.

---

5 The terms ‘language documentation’ and ‘documentary linguistics’ are both in current use and are often used interchangeably when referring to the subdiscipline of linguistics concerned with the interests outlined above. ‘Language documentation’ is also often used to refer to the material outcomes of documentary linguistic practices—that is, to the collections of resources assembled in documentary linguistics, rather than the overall enterprise (cf. Himmelman 2008: 346). Where possible, this study attempts to reserve the term ‘documentary linguistics’ to refer to the subdiscipline of linguistics and its associated theoretical and methodological concerns, and ‘language documentation’ to refer to the activities and products of research in this area.

The emergence of documentary linguistics as a distinct subdiscipline over the past twenty years arguably reflects both increased attention to the empirical foundations of contemporary linguistic research, as well as growing concerns over historically unprecedented levels of language loss around the globe and their implications for linguistic science (cf. Whalen 2004, Himmelmann 2008, Himmelmann 2012: 187). Indeed, the latter issue of language endangerment has often been raised in the documentary linguistic literature as having meaning both to the academic linguistic community and to affected communities of speakers and their descendants (cf. Hill 2002, Harrison 2007, Evans 2010, Mulder & Sellers 2010: 54). One consequence of this dual articulation of language loss has been considerable attention to the relationships between community-external and community-internal stakeholders in the development of documentation. This has contributed to extensive discussion of the role of collaboration in documentary linguistics and of models of research that attempt to recognize and incorporate the interests of both linguists and non-linguists in setting the direction of documentary activities (Dwyer 2006, Yamada 2007, Czaykowska-Higgins 2009, Leonard & Haynes 2010). Although recent debate has questioned the degree to which collaboration in the forms generally advocated in this literature is an ethical precondition to linguistic research in smaller and endangered language communities (cf. Crippen & Robinson 2013), there has nevertheless been a notable focus in documentary linguistics on such issues that have been decidedly less prominent in many other areas of mainstream linguistic research.<sup>6</sup>

This emphasis on collaboration in documentary linguistics also has consequences for the products of documentation. In contrast to research conducted with linguistic uses primarily in mind, language documentation is often undertaken with linguistic analysis as one anticipated application of the resulting collections among many, alongside other potential uses by academic and non-academic audiences in areas such as linguistic advocacy, education, and revitalization. Czaykowska-Higgins (2009) contrasts several models of documentary research contributing to such outcomes, distinguishing research *on* a language or community (where local interests in the

---

6 There has, however, been critical debate in anthropology and anthropological linguistics over the ethics of fieldwork practices throughout much of the late twentieth century and up to the present, although much of this discussion does not appear to have been taken up to the same extent in research on linguistic structure in neighbouring (sub)disciplines (cf. Himmelmann 2008: 338).

outcomes of research or their applications in the community may not be reflected substantially in either the design or conduct of research activities) from research *for* the community (where control over project goals and methods are still largely in the hands of linguistic researchers, but with some benefits of research intentionally shared with other stakeholders as a form of “giving back”) and research *with* the community (where all stakeholders are involved in the definition and refinement of the project at all stages in its development; see also Wolfram 1993 and Wolfram, Reaser & Vaughn 2008 for parallels in sociolinguistics). The intentional involvement of multiple stakeholders from the outset, even when differing perspectives on research exist, contributes to the kinds of reuse that documentary linguistics has among its aims, helping ensure that contributions to the documentary record are produced and represented in a manner acceptable to all parties involved.

This perspective on the multiple possible uses and interpretations of documentary records extends well beyond the more limited role afforded texts in the traditional ‘Boasian trilogy’ of grammar, text, and lexicon as a benchmark for comprehensive linguistic description, and has been influenced substantially by the advancement of digital technologies in linguistic science (Woodbury 2003, Whalen 2004, Evans & Dench 2006). Whereas earlier forms of linguistic description were limited by the means available for capturing linguistic observations, common digital technologies have rendered the production of richly annotated collections of audiovisual documentation commonplace—a development with significant consequences not only for conceptions of documentary linguistics, but also for the capacity of documentation-based projects to draw on such information in ways that would have been largely unimaginable even two decades ago (cf. Bird & Liberman 2001, Schultze-Berndt 2006).

In general, the documentary linguistic emphasis on developing and exploiting reusable, consistently annotated collections of observed linguistic behaviour and statements of metalinguistic knowledge blurs the lines between this discipline and parallel developments in other fields. Documentation-focused studies of German language enclaves in North America, such as the Texas German Dialect Project (Boas 2007, 2009), have explicitly sought to build bridges between the methodological recommendations of documentary linguistics and the theoretical concerns of research on German *Sprachinseln*. In the case of this project, the outcomes of a hybrid ‘documentary sociolinguistic’ approach have resulted not only in the

development of a singular resource on the history and culture of Texas German speakers accessible to community use, but also an opportunity to consider the “underlying dynamics of dialect contact and mixing, language contact, and language death” (Boas 2009: 100). Examples such as these suggest that the attention to the treatment of linguistic data that is central to documentary linguistics is not incompatible with the methods and research questions of other areas of linguistic inquiry. On the contrary, the individual concerns of different subdisciplines and other stakeholders motivate the compilation of different kinds of data, contributing through their specific interests to the breadth of documentation and to a realistic assessment of the practicality of current ‘good practice’ recommendations in documentary linguistics as a whole (Boas 2007). In a similar way, documentary linguistic perspectives may contribute a greater degree of critical attention to the theoretical, ethical, and technical issues raised in the context of research in the Saskatchewan Valley, and thus also encourage further reflection on the potential of collaborative approaches to investigating local linguistic variation.

#### 1.2.5 *Corpus linguistics*

The collections of primary data produced in documentary linguistic studies present an attractive target for corpus linguistics and associated quantitative approaches to linguistic analysis.<sup>7</sup> Despite sometimes substantial differences between documentary and corpus linguistics in their perspectives on linguistic data and the relationships between stakeholders involved in their development, the creation of documentation-based corpora provides another interpretive layer over documentary collections, opening their contents to additional forms of analysis (cf. Cox 2011b).<sup>8</sup> Indeed, corpus-based studies of linguistic variation are increasingly common in all of the subdisciplines surveyed above, in part due to a shared emphasis on the development of corpora as a significant component of research practice (e.g., Bauer 2002, Meyer

---

7 For the present purposes, the open debate as to the proper status of corpus linguistics as a distinct subdiscipline of linguistics (e.g., Leech 1992, Teubert 2001) or as a set of methodologies for linguistic analysis that are applicable to a range of fields (e.g., McEnery, Xiao & Tono 2006) is not critical to this research, which seeks to draw on corpus-based work from both perspectives without requiring a definitive answer to this question.

8 This study distinguishes between documentary linguistic *collections*, consisting of primary data assembled through language documentation, and corpus linguistic *corpora*, which may involving selective sampling and additional interpretation of such primary data; see Cox (2011b: 261) for discussion.



2004, Kretzschmar et al. 2006, Moisl 2009, Szmrecsanyi 2011, i.a.). Even where differences are noted in the forms of language to which each discipline typically attends, it bears noting that these areas are already in contact with one another in this respect, and that much research into language variation is already based on the use of corpora and corpus-based analytical methods.

Perspectives from corpus linguistics are also relevant in matters of corpus construction. Concerns similar to those expressed in variationist sociolinguistics about representativeness are well represented among corpus linguists, as well. Biber (1993) raises a series of issues in the definition of populations and texts for the development of balanced corpora that are comparable to the concerns of Sankoff (2005) in sociolinguistics. Crowdy (1993) also advances similar principles of representativeness for spoken corpus design, describing procedures used in the development of the spoken component of the British National Corpus to achieve even representation of particular contexts of use and the overall demographics of the English-speaking population of the United Kingdom. In addition to more abstract concerns over the design and planning of corpora, there is a considerable literature on corpus construction procedures in practice (e.g., Wynne 2005) that provides valuable counterbalance to theoretical arguments over the ideal structure of linguistic corpora.<sup>9</sup>

Despite these apparent compatibilities, one should also recognize that corpus linguistics and other areas of linguistic research also show clear divergence. Kendall (2011) argues that contemporary corpus linguistics shows a marked bias towards majority, standard varieties, in contrast to the non-standard varieties of larger languages on which variationist sociolinguistic research has more often concentrated. Although notable exceptions to this pattern exist (e.g., in the multilingual research agenda for corpus linguistics proposed by McEnery & Ostler 2000, or the dialect corpora described in Johannessen 2011 and Szmrecsanyi 2011), the significant proportion of corpus linguistic research that has been devoted to majority languages stands in contrast to much work in documentary linguistics, dialectology, and dialectometry. Similarly, the tendency of corpus linguistics to concentrate on developing corpora from large collections of existing linguistic material has also contributed to the relative rarity of corpora of spoken

---

<sup>9</sup> Similar reports on the practice of corpus collection, annotation, and management are also found in variationist sociolinguistics, although somewhat less prominently than in corpus linguistics; see Poplack (1989), Tagliamonte (2006), and Childs, van Herk & Thorburn (2011).

language (cf. Newman 2008). This presents a significant cleft between the corpora typical in mainstream corpus linguistics, with its tendency towards extremely large collections of standard varieties of majority languages drawn primarily from written media and the corpora of other linguistic subdisciplines, which are often considerably smaller, but provide greater proportional representation of spoken, non-standard, and demographically less prominent varieties and languages.

Such differences notwithstanding, corpus linguistic perspectives on linguistic variation may contribute to an understanding of Mennonite Plautdietsch in the Saskatchewan Valley in several ways. Contemporary corpus linguistic studies demonstrate the application of a particularly extensive range of quantitative methods to observations of language use. This methodological breadth is well aligned with recent work in sociolinguistics and dialectometry, where statistical methods are already well established and continue to expand to include new techniques.<sup>10</sup> Similar methods might be adopted to come to a clearer understanding of the linguistic situation in the Saskatchewan Valley. Corpus linguistics also provides theoretical and practical recommendations on the compilation of corpora, despite an apparent bias towards standardized, written forms of majority languages. In this way, it offers a counterpoint to similar recommendations on developing collections of linguistic records in documentary linguistics, albeit with the benefit of more extensive experience with the later application of the resulting corpora to problems of linguistic analysis, an area which is only now beginning to be explored in depth in documentary linguistics (cf. Seifart et al. 2012).

### *1.2.6 Summary*

Although by no means exhaustive, the preceding sections suggest several areas in which the present research might benefit from the perspectives of different linguistic subdisciplines:

- Variationist sociolinguistics' commitment to understanding the social embedding of variation underscores the importance of considering sociodemographic factors in the

---

<sup>10</sup> For instance, see Tagliamonte & Baayen (2012) for applications of random forests, conditional inference trees, and mixed-effects modelling to problems in variationist sociolinguistics; and Wieling, Nerbonne & Baayen (2011) on generalized additive models in dialectometry. Further consideration of these techniques is given in Chapter 6.

Saskatchewan Valley as potentially relevant predictors of local linguistic differentiation. Likewise, the frequent application of quantitative methods in sociolinguistic research to spoken language corpora in modelling linguistic variation suggests that similar methods might also be applied productively in the present research;

- By comparison, dialectology and dialectometry place greater emphasis on the investigation of geography as a potential predictor of the distribution of linguistic variation across speech communities. This focus on geolinguistic variation is reflected in the multivariate comparisons made in this tradition between geographically disparate communities, with dialectometry further attempting to establish connections between dialectological survey data and quantitative statistical methods. Although the latter approaches have focused primarily on analyses of large-scale dialect surveys in historically well-established speech communities, it may be possible to adapt these methods to smaller samples of speech in more sociolinguistically diverse groups, affording both the benefits of dialectometric methods and an opportunity to extend and evaluate their use in other linguistic contexts (cf. Stanford 2012 for a recent example of dialectometric methods being applied in the context of a small, clan-based society);
- Documentary linguistics shares with the aforementioned disciplines a concentration on reusable collections of linguistic data, but gives greater attention to theorizing the compilation, preservation, and use of such resources as language documentation. This emphasis on the treatment of raw and primary linguistic data (in the sense of Himmelmann 2012) throughout their collection, annotation, and application stems in part from engagement with issues of language endangerment and loss, prompting documentary linguistic concerns over both the empirical adequacy of existing linguistic records for many smaller languages and the long-term accessibility and usability of these resources for future audiences. These same considerations are also reflected in active documentary linguistic debate over ethical considerations in linguistic fieldwork, and particularly the role of collaboration in language documentation, presenting points which merit discussion in planning documentation of Mennonite Plautdietsch;
- Finally, corpus linguistics complements the above disciplines in the breadth of methodological approaches it contributes to the analysis of linguistic data. Corpus-

based, quantitative methods might contribute to understanding the structure of linguistic variation that the present study aims to explore, and also assist in bridging between permanent language documentation on the one hand and descriptive applications on the other.

Importantly, these perspectives are seldom inherently at odds with one another. Despite differences in emphasis and interpretation, one finds commonalities in the importance these fields place on corpora of observed linguistic behaviour and in their preference for analytical methods that assume these records as their basis. Clearly, these are not the only linguistic perspectives possible on the situation of the Saskatchewan Valley, nor is this summary meant to suggest that each of these disciplines shows absolute uniformity even in these identified points of intersection. Nevertheless, the observations made here present a reasonably coherent set of desiderata for the investigation of linguistic differentiation in the present communities. The application of these perspectives to the present study will be explored further in Chapter 4, after considering the history of the Russian Mennonites and prior studies of their linguistic practices in the following chapters.

## **2 Mennonite history and the Saskatchewan Valley**

### *2.1 Introduction*

Understanding the linguistic practices of the Saskatchewan Valley Mennonite communities requires consideration of the broader historical and social context in which they are embedded. As elsewhere in the Russian Mennonite diaspora, extensive histories of contact with and isolation from other Mennonite and non-Mennonite groups have contributed to the constitution of the Saskatchewan Valley Mennonite communities as they exist today, and one might expect this to be reflected in the current linguistic practices of these communities, as well. These histories of contact in and across periods of settlement provide valuable points of relation with communities in the larger Russian Mennonite diaspora, allowing comparisons to be drawn between the Saskatchewan Valley and other Plautdietsch-speaking communities, drawing attention to linguistic phenomena reported to vary among Plautdietsch varieties that may otherwise be overlooked, and encouraging consideration of how linguistic differences between diasporic groups may relate to their individual histories of separation and contact.

With this in mind, this chapter describes the historical paths that brought Russian Mennonites to the Saskatchewan Valley in the late nineteenth and early twentieth centuries. Beginning with their emergence in the Radical Reformation of the early sixteenth century, the following sections detail the flight of many early Mennonite Anabaptists to areas of northern Poland (§2.2), as well as subsequent migrations to Ukraine (§2.3), the USA and Canada (§2.4), and later movements which resulted in the settlements in the Saskatchewan Valley that are at the heart of this study (§2.5, §2.6). The experiences of these groups after their immigration to Saskatchewan are described in Section 2.6, concluding with a brief discussion of the current linguistic situation in the region.

### *2.2 Mennonite origins and migration, 1525–1787*

The Mennonites represent an Anabaptist Christian denomination that emerged in central Europe as part of the Radical Reformation (Bainton 1952, Dyck 1993). These groups' insistence as a matter of core doctrine upon adult baptism, non-violence, and a refusal to swear oaths was perceived as a threat by states, which commonly relied on infant baptismal records for purposes of taxation and military conscription. This led to the severe persecution of early Anabaptists at

the hands of both Protestant and Catholic authorities, although ultimately failing to prevent the rapid spread of Anabaptism from its origins in central Europe in the first half of the sixteenth century. Menno Simons (1496–1561), a Frisian Catholic priest, left his clerical office in 1536 to join with local Anabaptists, becoming an influential itinerant leader among several pacifist Anabaptist groups to which his name would later be applied (cf. Reger & Plett 2001: 15).

Heavy persecution during this period led to significant movement of Anabaptist populations, with some groups from the northern European lowlands and central highlands congregating in northwestern Europe, where religious persecution was less pronounced. An eventual shift in the local political climate ended the tolerance that had temporarily been extended to Anabaptists in East Friesland, and led to widespread eastward migration of the refugee group that had assembled there (Epp 1993: 50–51). This flight from the northwest led to many areas, among them the northern region of present-day Poland, where a policy of religious tolerance had been enacted by the Prussian states in an effort to attract settlers to the Vistula River delta (Schapansky 2006: 66). Despite strong opposition on the part of local trade guilds and the Lutheran church, as well as initial resistance to Mennonite settlement inside of free cities, significant Mennonite migration into the region followed.

Mennonites entering the Vistula delta as refugees from northwestern Europe in the sixteenth century were neither ethnically nor linguistically uniform. As both Epp (1993) and Schapansky (2006) stress, Mennonites of this group consisted not only of Frisians and Saxons from East Friesland and surrounding areas, but also contained a substantial contingent of refugees from Flanders, as well as a smaller but still significant number of central Europeans. Groups within this population were distinguished not only by language and ethnicity, but also by denominational divisions initiated in 1566, which split these Mennonites into ‘Flemish’ and ‘Frisian’ factions that persisted in descendant communities for several centuries (Schapansky 2006: 44–45). While the names of these factions reflect their initial division along ethnic lines, significant crossover between members of both groups in later years reflects the importance of ideological differences in distinguishing them. Members of the Frisian group generally favoured a greater division between spiritual and secular life and placed somewhat more emphasis on individual freedom in decision-making, while the Flemish group stressed the close integration of spiritual and secular life and an orientation towards viewing the actions of the individual in the



Figure 2. Vistula delta region.

context of the larger religious community (Schapansky 2006: 46, Epp 1993: 71). Further schisms would follow in both groups, but a sense of general differentiation between ‘Frisian’ and ‘Flemish’ populations was still evident in later migrations (cf. Quiring 1928: 42–45).

Mennonites entering the Vistula delta at this time arrived in a linguistic landscape that was no less diverse than their own internal linguistic and ethnic-denominational composition. Both Ziesemer (1924: 125) and Epp (1993: 67–68) report nine distinct Low German dialects in the region, as well as several neighbouring High and Middle German varieties. This established geographical dialect variation is critical to several hypotheses concerning linguistic differences between descendant Mennonite communities (cf. §3.2.5). Such accounts commonly distinguish three zones in the area of northern Poland shown in Figure 2 in which most Mennonites came to

settle: (a) a narrow spit of land (Standard German: *Nehrung*) north of the Elbląg branch of the Vistula, extending from the city of Gdańsk (formerly Danzig) to the northern edge of the Vistula Lagoon; (b) several large, lowland river islands (*Werder*) to the south of the Elbląg Vistula; and (c) an extensive, low-lying valley region (*Niederung*) further south. While early Mennonite settlement was concentrated primarily around major cities in the northern delta, particularly in the areas of Gdańsk, Elbląg (formerly Elbing), and Malbork (formerly Marienburg), a significant Mennonite presence later extended into surrounding areas, thus encompassing a highly heterogeneous dialect landscape (Epp 1993: 68).

Although these varieties of Low German were significant features of the local linguistic landscape for newcomers to the region, Mennonite refugees were likely to have had economic, religious, and personal motivations for maintaining or acquiring some ability in Dutch or High German, as well. As Schapansky (2006: 73) observes, Dutch was of particular economic importance in cities such as Gdańsk as a result of strong maritime trade connections between West Prussia and the Netherlands, particularly in the midst of the decline of the Hanseatic League and the waning influence of its earlier Low German lingua franca. These commercial ties, accompanied by efforts throughout the eighteenth century to maintain contact between Mennonite religious communities (*Gemeinden*) in West Prussia and the Netherlands, fostered the retention of Dutch in these areas. By comparison, Mennonites situated farther south in the Vistula valley, predominantly belonging to Frisian denominations, were at once less exposed to the economic benefits of competency in Dutch and in closer contact with neighbouring High German-speaking populations. Thus, while there is some historical evidence in church records to suggest that a transition to High German as the language of worship among the Frisian *Gemeinden* concentrated in the Vistula valley had taken place by the mid-seventeenth century, the predominantly Flemish *Gemeinden* in the northern delta maintained Dutch as their language of church and written communication for another full century (Schapansky 2006: 73).

While Mennonites in Poland thus maintained either Dutch or High German as their language of written communication and worship, they nevertheless came to adopt local varieties of Low German as their language of daily life. The result of this shift in vernacular, coupled with the intentional maintenance of a distinct written variety, was a stable form of societal bilingualism that persisted in similar configurations across subsequent Mennonite migrations for



several centuries. This sociolinguistic arrangement later became entrenched as an important cultural practice for the ethnic and religious identity of many descendant Mennonite communities, and would present another significant catalyst for later migrations (cf. Hedges 1996, Warkentin 2010).

While the Mennonite population in northern Poland continued to expand throughout the seventeenth and eighteenth centuries, both internal and external political and economic pressures at the end of the eighteenth century encouraged many Mennonites to consider the offer of Catherine II to settle lands in present-day Ukraine that had recently been acquired from the Ottoman Empire. Through negotiations with delegates of the Russian state, Prussian Mennonites were able to secure a series of concessions in 1786 (later referred to as the *Privilegium*) which confirmed Mennonite religious and educational autonomy, land rights, and exemptions from taxes and military service in the Russian Empire. The first migration from northern Poland to Ukraine took place in 1787, with 228 Mennonite families establishing the first Mennonite settlement in the Russian Empire, the Chortitza Colony, in 1789.

### 2.3 *Mennonites in Ukraine, 1787–1873*

Mennonite migration from Prussia to the Chortitza or ‘Old’ Colony continued well into the early nineteenth century, rapidly establishing a significant Mennonite presence in the area. The extent of immigration soon necessitated the formation of a second major Mennonite settlement, the Molochnaya or ‘New’ Colony, established in 1804 by 355 families some 120 kilometres east of the Chortitza Colony (Epp 1993: 77, Schroeder & Huebert 1996: 124). As with the Chortitza Colony, this new settlement quickly grew through the influx of Prussian Mennonites until 1806, when immigration largely stalled as a result of the Napoleonic Wars. Whereas Mennonite immigrants to both colonies prior to the wars were largely members of more ‘traditionalist’ Flemish Mennonite *Gemeinden* in the Vistula delta, the later, post-war group showed much stronger representation of Frisian Mennonites who had settled further south in the Vistula valley, and whose reportedly more accommodating attitudes towards economic and social integration had perhaps provided less motivation for emigration from Prussia before the intervention of conflict in the region (Schapansky 2006: 143–145).

After weathering initial hardships in the establishment of both colonies, a period of

significant expansion followed, fuelled both by internal population growth and post-war immigration from Prussia which continued into the 1830s. This expansion soon brought about further land shortages, which were of immediate concern for communities with a primarily agricultural economic basis. This motivated the creation of daughter colonies elsewhere in the Russian Empire, with Bergthal and Fürstenland established as daughter colonies of Chortitza in 1836 and 1864, respectively (Doell 1987: 3, Guenter et al. 1995: 9, Reger & Plett 2001: 333, 436), and still others as daughter colonies of Molochnaya. Although connections were often maintained between historically related Mennonite settlements, both the mother colonies and their daughter colonies were essentially closed and largely self-sustaining from the perspective of neighbouring populations, as was encouraged by contemporary Russian colonial policy. One consequence of this model of autonomous, ethnically homogeneous agrarian settlement was the gradual emergence of a distinct Russian Mennonite ethnic and linguistic identity. This first instance of geographical separation between Mennonites and the larger northern continental Germanic dialect continuum, combined with trade contact with speakers of Ukrainian and Russian, contributed to linguistic innovations which represent perhaps the first independent developments of distinctly Mennonite varieties of Plautdietsch (cf. Quiring 1928: 108ff, Wiens 1957, Kaufmann 2003a).

Mennonite settlement in Imperial Russia continued largely unabated until the 1860s, when the state began to exert pressure on Mennonite communities to introduce compulsory Russian language education into colonial schools and to participate in military service (Epp 1962: 25–26, Epp 1993: 82–83). Combined with the problem of recurring land shortages, concerns over the revocation of the Mennonite *Privilegium* in 1871 prompted the migration of almost one third of the Russian Mennonite population to North America between 1874 and 1880. Of the approximately 17,000 Mennonites who left for North America during this period, some 10,000 (predominantly from the Molochnaya Colony) emigrated to the midwestern United States, settling in Kansas, Nebraska, Oklahoma, Minnesota, and the Dakotas. The remainder immigrated to southern Manitoba, where exemption from military service and a degree of freedom in educational matters had been assured by representatives of the Dominion of Canada (Dyck 1993: 206–207, Epp 1993: 84).<sup>11</sup> This latter group was predominantly, though not

---

<sup>11</sup> These groups were not alone among Mennonites in Canada, or even North America. These Russian Mennonites

exclusively, from the Chortitza Colony and its daughter settlements, including approximately 3,000 Mennonites from the Chortitza Colony itself, 3,000 from Bergthal (representing essentially the entire colony), 1,000 from Fürstenland (again representing almost the entire colony), and a smaller contingent of about 700 Mennonites from the Kleine Gemeinde denomination in the Molochnaya Colony (Doell 1987: 3, Plett 2001: 26, Plett & Reger 2001: 537).

#### 2.4 *Mennonites in North America, 1874–1890*

The first Mennonites arriving in Manitoba settled on eight townships east of the Red River that had been designated by the Dominion government for exclusive Mennonite settlement. These reserved lands were settled primarily by the Bergthaler and Kleine Gemeinde denominations, who established the first villages there in 1874 (Schroeder & Huebert 1996: 142). Both the Fürstenland and Chortitza groups, who arrived the following year and merged their denominations to form the Reinländer *Gemeinde*, sought lands between the Red River and the Pembina Hills further west of the existing Mennonite settlements. These lands were eventually granted as part of an additional eighteen townships set aside by the Dominion government for Mennonite settlement in 1876. This division into two primary settlements on opposite sides of the Red River, the East Reserve and West Reserve, became the basis of Russian Mennonite colonization in southern Manitoba.

In these reserves, Mennonite immigrants sought to re-establish a pattern of traditional, village-based settlement that had developed during their time in northern Poland and Ukraine. This *Strassendorf* model of settlement involved linear, single-street villages consisting of approximately twenty families, with community structures (typically a church and a school) often located towards the centre of the village. Participant families combined their lands into a single holding, which was then subdivided into both common pasture and individual, equally sized parcels of land for each family to cultivate (Dawson 1936, Friesen 1975). Typically established through the planning of individual *Gemeinden*, such villages relied on groups of families bound together not only by common religious and ethnic identity, but also by their contributions to the maintenance of shared infrastructure and social institutions that underlay the

---

were preceded in both the USA and Canada by long-established Swiss Mennonite communities, which, despite their common Anabaptist religious heritage, were linguistically and culturally distinct (Kloss 1989).

autonomy and self-sufficiency of these settlements. The re-establishment of Mennonite institutions charged with maintaining social security, including fire insurance (*Brandornung*) and a social assistance agency for widows and orphans (*Waisenamt*), likewise encouraged the development of autonomous networks of support anchored in the *Gemeinde* and limited the relative importance of sources of economic or social aid outside of the Mennonite community.

The result of the transplantation of these traditional settlements and social institutions in the Canadian context was dense, multiplex networks of areal, denominational, and kinship ties in and between largely self-sustaining communities, each having only minimal systemic reliance on outside groups for needs beyond agricultural commerce and basic trade. The limited overall participation of community members in non-Mennonite society (and, to some extent, even with members of other *Gemeinden* not directly linked by either kinship or geography) followed partly from the degree of community coherence to which these structures contributed.

### 2.5 *Mennonites in the Saskatchewan Valley, 1890–1921*

Rapid population growth in the first twenty years of Mennonite settlement in Manitoba again brought issues of landlessness to the fore, providing substantial motivation for westward movement into the Northwest Territories. The first major westward migration into the District of Saskatchewan from southern Manitoba took place in 1891, when eleven Reinländer Mennonite families, returning from a failed attempt at settlement near Gleichen, Alberta, were convinced by a land agent to take up homesteads near the town of Rosthern, Saskatchewan. Although not an ‘official’ settlement, this small group was soon joined by another 27 families in the spring of 1892, some immigrating directly from Ukraine, others from southern Manitoba (Guenter et al. 1995: 21, Doell 1987: 6). This settlement near Rosthern represented the first large community of Russian Mennonites to establish permanent homes in the District of Saskatchewan (Ens in Guenter et al. 1995: iii).

Even with this trickle of unorganized westward migration easing some of the economic pressures that persistent land shortages in Manitoba presented for local Mennonite communities, such concerns prompted negotiations between representatives of the Reinländer *Gemeinde* and the Canadian Minister of the Interior. These discussions concluded in 1895 with the establishment of the Hague-Osler Mennonite Reserve, four townships in the Saskatchewan

Valley between the North and South Saskatchewan rivers intended exclusively for Mennonite settlement (Guenter et al. 1995: 26). These lands were rapidly settled by Reinländer Mennonites from Manitoba, with the first 90 settlers arriving in the spring of 1895, followed by an additional 218 families between 1898 and 1904.<sup>12</sup> The extent of this immigration soon necessitated further negotiations with the government, leading in 1898 to three expansions of the reserve to include lands to the north and south of the core allotment (Guenter et al. 1995: 26–28), as well as the establishment of a second colony near Swift Current, Saskatchewan, in 1904 (Doell 2001).

Reinländer Mennonites entering the Hague-Osler Reserve during the first two decades of its existence established a considerable number of settlements, including not only seventeen traditional *Strassendorf* villages, but also several homestead-based ‘four-corner hamlets’ and more sparsely populated rural districts consisting primarily of individual homesteads (Friesen 1975, Guenter et al. 1995: 27). Appendix A summarizes the permanent Mennonite settlements founded in the Saskatchewan Valley before 1909. This rapid influx of Mennonite settlers and the expansion of railway connections into the area provided support for both Mennonite and non-Mennonite businesses to develop in nearby towns and villages. This contributed directly to the development and incorporation of the communities of Hague (1903), Osler (1904), Warman (1905), and others in the Saskatchewan Valley until the end of second decade of the twentieth century. These predominantly non-Mennonite communities are profiled in Appendix B.

Although there was thus heavy representation of Reinländer Mennonites in the areas negotiated as the Hague-Osler Mennonite Reserve proper, this was far from the only Mennonite group in the Saskatchewan Valley. Bergthaler Mennonites, predominantly the descendants of immigrants from the Bergthal Colony in Ukraine, were present in Saskatchewan since 1892, arriving from Manitoba with the second wave of Reinländer settlers in the Rosthern area and establishing the Bergthaler Mennonite Church in 1893 (Doell 1987: 6, 15). Most Bergthaler groups were situated at the edges of the Hague-Osler Reserve, or dispersed in smaller groups throughout the Saskatchewan Valley (Doell 1987: 13, 15). Alongside the Reinländer

---

<sup>12</sup> Assuming roughly six people per family (drawing this estimate from the report of 90 individuals among the 14 families from Manitoba who settled in the Hague-Osler Reserve in 1895; Guenter et al. 1995: 26), one can conservatively estimate that more than 1,300 Reinländer Mennonites had established their residence in central Saskatchewan within the first decade of the reserve's existence.

Mennonites, the Bergthaler Mennonites represented demographically one of the most significant communities in early Saskatchewan Valley Mennonite settlement, with the Saskatchewan *Gemeinde* counting 440 baptized members and 1,000 total souls by 1905 (Doell 1987: 114).

At the same time, a smaller group of five Mennonite families also settled near Rosthern in 1891, representing members of the Rosenorter *Gemeinde* who emigrating directly from the area of Tiegenhof, Prussia, to the Saskatchewan Valley (Brednich 1977: 17). Unlike the Reinländer and Bergthaler denominations, whose memberships consisted primarily of families who had remained adherents of these *Gemeinden* for several generations, members of the local Rosenorter *Gemeinde* stemmed both from the immigration of adherents from Prussia, Ukraine, and the USA, as well as from transfers in membership from other Mennonite denominations in Saskatchewan and Manitoba (cf. Guenter et al. 1995: 615). Likewise, not all members of the Prussian contingent in the Rosenorter *Gemeinde* maintained Plautdietsch as their language of the home, having transitioned to the daily use of High German prior to their immigration to Canada (cf. Quiring 1928: 47, Tolksdorf 1985: 323–324). Although in the extreme minority among these groups (as in other Mennonite settlements in the Americas; see Rohkohl 1993: 107–108 on the experience of Polish Mennonite immigrants to Paraguay), distinctive ‘Polish’ varieties of Plautdietsch are attested, presenting an additional point of diversity on the local linguistic landscape.<sup>13</sup> By comparison, the American contingent in the Rosenorter *Gemeinde* largely comprised groups with historical ties to the Molochnaya Colony who settled in the American mid-west and whose northward migration into Canada was generally on the more modest scale of individual and extended families, rather than of entire communities leaving en masse (cf. Buchheit 1982, 1988, Keel 2006, Doell n.d.). Unlike their Polish coreligionists, these Rosenorter Mennonites from the USA generally maintained the traditional Plautdietsch-High German bilingualism (Epp 1972).

Some American Mennonites arriving as homesteaders in the Saskatchewan Valley at this time were also adherents of the Mennonite Brethren *Gemeinde* (Epp, Epp & Thiessen 2009). As a denomination with roots in a religious revival movement which took place in the Molochnaya

---

<sup>13</sup> A rare example of Polish Plautdietsch in the Saskatchewan Valley is found in interview BRE 6 (2), conducted by Rolf W. Brednich with Ernest A. Jeschke in 1977 and now held in the collections of the Canadian Museum of Civilization (Brednich 1977; CMC Archives Sound Recordings VII-C-104).

Colony in the 1860s, the Mennonite Brethren were particularly prominent among Russian Mennonites immigrants to the USA with historical ties to that settlement. Together with Mennonite Brethren immigrants arriving from Ukraine, this community established its first Saskatchewan church in 1901 in Bruderfeld, southwest of Waldheim. These congregations were strengthened substantially by later Mennonite Brethren immigration from the USA into the Hepburn area around 1910 and by connections to Mennonite Brethren communities in southern Manitoba, which already counted 1,266 members by 1887 (Guenter 1981: 25, 624).<sup>14</sup>

Although each of these Mennonite groups—Reinländer, Bergthaler, Rosenorter, and Mennonite Brethren—was thus distinct, settlement in Saskatchewan often brought individuals from these denominations into close contact with one another. In the early settlement period, individuals who lived some distance away from their nearest affiliated church would often join together with members of other denominations for church services, thus encouraging interaction among these groups (cf. Doell 1987: 9, 11). Similarly, among the more conservative denominations, private schools were often organized by Reinländer Mennonite communities, but often had both Bergthaler and Reinländer children in attendance (Doell 1987: 22, Guenter 1981: 32), presenting another important instance of persistent inter-group contact.

While such mutually supportive relationships between denominations and their members provided critical support for the establishment of Mennonite settlements in the Saskatchewan Valley, not all elements of the traditional settlement model were instituted as planned. The *Strassendorf* model of village-based colonization faced competition from the individual homesteads offered by government colonization agencies in western Canada. While amendments to the Homestead Act in 1907 effectively prevented the incorporation of further single-street villages as governmentally-recognized hamlets (as was required to dispense with the legal requirement that individuals maintain residency directly on their homesteads), this mode of settlement faced serious internal challenges even before this shift in governmental policy. In

---

14 While historical ties to Prussia and the Molochnaya Colony immediately distinguished many Rosenorter and Mennonite Brethren Mennonites from their Bergthaler and Reinländer neighbours, it is also important to note that the former communities were not homogeneous with respect to the immigration history of their members. Both transfers of membership from other denominations and active programs of evangelization (sometimes focused on members of other Mennonite *Gemeinden*; see Plett 2000, 2001) contributed to a considerable range of denominational and emigrational backgrounds in these denominations.

both Saskatchewan and Manitoba, early Mennonite settlements rapidly transitioned from common holdings to individual homesteads, placing considerable strain on *Gemeinde*-administered support systems that required almost total participation to remain viable.<sup>15</sup> This did not result in the dissolution of the villages *per se*—single-street settlements continue to exist on the Canadian Prairies in a geographically recognizable form, as noted in Appendix A—but rather contributed to a radical restructuring of their role in maintaining social and economic coherence in Mennonite communities and the families that constituted them.

While these shifts in settlement pattern had notable consequences for relations within and between Mennonite communities in Saskatchewan, greater challenges were to follow in matters of education. As was noted above, Mennonite schools were traditionally established under the direction of individual *Gemeinden*, with schools in the Hague-Osler Reserve often being shared between Bergthaler and Reinländer *Gemeinden* (albeit generally under Reinländer organization and leadership; Guenter 1981: 32, Doell 1987: 22). These schools provided universal education in Standard German centred around basic literacy, numeracy, and writing skills, and emphasized a standardized progression from basic texts in each of these domains to more specialized, often religious literature in the form of primers (*Fibels*), the Bible, and the 1778 Elbing catechism. This progression reflected a broader conception of education as a critical element in the formation and maintenance of the community, both preparing children for eventual adult membership in the *Gemeinde* and reinforcing the linguistic and cultural practices that supported the larger religious order (*Ordnung*; cf. Redekop 1969, Hedges 1996).

As schools were thus administered through the *Gemeinden*, there were at times educational consequences for the children of members of the community who faced church discipline. The protests of several former Saskatchewan Reinländer Mennonites, concerned in part with the consequences of their children not attending a *Gemeinde*-run school, led in 1908 to a Royal Commission of Inquiry into the state of Mennonite private schools in Saskatchewan Valley (Doell 2001: 144). Although the Reinländer *Gemeinde* leadership was ultimately able to

---

<sup>15</sup> Although motivations for this transition are reported to have varied considerably, ranging from concerns over the quality of commonly held lands to the effects of individual changes in denominational affiliation (cf. Plett 2001), as Friesen (1975) notes, individual homestead ownership also furnished the large, unsubdivided tracts of land required for entry into the large-scale commercial grain market, providing additional financial incentive to shift away from the *Strassendorf* system.



negotiate the continuation of these schools under local control, this incident placed the Mennonite educational system under increased scrutiny. This attention was only heightened with the advent of the First World War and a concomitant rise in British nationalist, anti-German, and anti-pacifist rhetoric in public discourse. This further intensified with Mennonite exemption from military service throughout the war, even after the enactment of the 1917 Conscription Act (Plett 2001: 17, Doell 1987: 22). The dominant political climate and a general emphasis in western Canada on providing standardized access to education in the British tradition contributed to the passing of School Attendance Acts in both Manitoba (1916) and Saskatchewan (1917). This legislation instituted mandatory attendance in English-language, provincially administered schools for all school-age children. This resulted in the forced closure of the Reinländer Mennonite schools in the Saskatchewan Valley and the establishment of English-language public schools on Mennonite reserve lands in 1919, in some cases on properties expropriated from Mennonite owners unwilling to sell them for this purpose (Ens 1994: 138).<sup>16</sup>

The consequences of this legislation prompted a moment of crisis for Reinländer and Bergthaler Mennonite communities in Manitoba and Saskatchewan. Earlier assurances of educational autonomy were dismissed as being outside of the legal purview of the federal government that offered them, with education falling under provincial authority in the Canadian system. The seeming abrogation of freedoms granted by the Canadian government to Mennonites was viewed in many communities as a significant challenge to their religious and social foundations, undermining not only their autonomy in decisions pertaining to education, but also in the ability of the *Gemeinden* to remain faithful to their understanding of their calling to be a ‘people apart’—in the world, but not of it. Despite formal appeals to the government for inspections of schools and an independent assessment of the Mennonite educational system and its outcomes, no such concessions were granted. Beginning in Saskatchewan in the spring of

---

16 The resistance of Saskatchewan Valley Reinländer Mennonites to the imposition of provincial schools is evidenced not only in their refusal to sell lands for the construction of these facilities and their refusal to pay governmental fines for non-attendance, but also in the actual relocation of entire families from their homesteads to locations beyond the bounds of new school districts. Such families (in some cases, living in groups of up to fifteen families in extremely limited housing on a single quarter-section) relied on clauses in the School Attendance Act which provided an exemption from mandatory attendance for children who lived beyond a certain distance from the nearest public school (Leonard Doell, 2011, p.c.).

1918, fines were imposed on individual Reinländer and Bergthaler families who continued to send their children to Mennonite private schools that did not comply with the new legislation.<sup>17</sup> This proceeded to the extent that officials were warned that Mennonite families were being “reduced to destitution through the fines being imposed upon them” (Ens 1994: 146). In cases where fines could not be paid, property was seized and, in some cases, individuals imprisoned (Ens 1994: 139). Continued prosecution of infractions of the School Attendance Act followed over the next decade, with some 2,346 Mennonites in the Saskatchewan Valley paying fines totalling over \$20,000 and twelve individuals being jailed in 1920–1921 alone (Ens 1994: 148).

While the Bergthaler Mennonites were also resistant to the imposition of governmental schools, the Reinländer *Gemeinde* was somewhat firmer in its insistence on non-compliance, instructing members not to pay fines and even threatening some members with excommunication who sent their children to English-language schools. With no resolution in sight, the Reinländer pursued plans for emigration as early as 1919, with the Bergthaler exploring emigration to South America in the years immediately following. This ultimately resulted in the emigration of between one quarter and one third of the Saskatchewan Reinländer Mennonites to northern Mexico beginning in 1924 (Epp 1982: 119). An additional 1,763 Mennonites emigrated to the Paraguayan Chaco in 1926, with 195 Bergthaler Mennonites from the Saskatchewan Valley participating in this move, albeit with substantial return migration (Doell 1987: 30).

## 2.6 *Mennonites in the Saskatchewan Valley, 1922–present*

The consequences of Reinländer and Bergthaler mass emigration from the Saskatchewan Valley were considerable, both for the overall demographic constitution of the region and for the cultural and linguistic developments that would follow. Epp (1982: 124) comments that these migrations “stunned the reserves in Manitoba and Saskatchewan and permanently altered the socio-religious complexion of these areas,” with a significant reduction in the proportion of these communities relative to other Mennonite groups in the Saskatchewan Valley. Those Reinländer Mennonites who did not join in the move to Mexico, now greatly reduced in number and without

---

<sup>17</sup> The installation of provincial schools bearing the names of prominent battle sites in the First World War—Passchendaele, Venice, Pembroke, and Renfrew—in pacifist communities has been suggested by some historians to have been met with further disapprobation on the part of affected Mennonite groups (Ens 1994: 134–135).

formal denominational connections to the new Mexican settlements, eventually reincorporated in 1930 as the Old Colony Church, but still faced governmental fines for non-compliance with the School Attendance Act. The extent of this migration resulted in the relocation of entire villages, opening up lands in the core Mennonite settlement area for non-Mennonite use.

At the same time, political developments following the Russian Revolution introduced a period of social upheaval for those Mennonites who had remained in the Russian Empire. The conclusion of the First World War and the events of the Russian Revolution destabilized southern Ukraine, leading to a period of anarchy during the Russian Civil War (1917–1922). These years saw Mennonite villages in southern Ukraine occupied at times by the Red Army and by Makhnovist anarchists, with armed forces seizing property, conducting executions, and committing other acts of violence. This period was followed almost immediately by outbreaks of typhus, tuberculosis, and widespread famine. These circumstances prompted a second mass migration to the Americas beginning in 1922, with some 20,000 Mennonites arriving in Canada as refugees over the next seven years, and over 600 in Saskatchewan in 1923 alone (Epp 1962, Dyck 1993: 188, Guenter et al. 1995: 344).

The concurrence of Reinländer-Bergthaler emigration to Latin America and post-Revolution Mennonite immigration to western Canada radically altered the settlement landscape of the Saskatchewan Valley. Lands offered for sale by members of outbound groups were often purchased by incoming refugees, particularly in the western half of the Hague-Osler reserve. Formerly heterogeneous villages were no longer enclaves of a single *Gemeinde*; rather, neighbours were now sometimes of different Mennonite denominations or non-Mennonites.<sup>18</sup> This period of initial contact between these two groups also brought to fore differences in experience and attitude between so-called *Russländer* (post-1920 ‘Russian’ immigrants) and *Kanadier* (pre-1920 ‘Canadian’ immigrants) Mennonites. Considerable documentation exists of

---

18 Dawson (1936: 106) notes a similar effect of the migration to Mexico in southern Manitoba when “the sectarian ‘core’ of the Mennonite communities was removed and large tracts of land in the very centre of the old Reserves were left vacant.” These lands were eventually “taken up, partly by those liberal Mennonites who did not emigrate, and partly by new immigrants from Russia who were even more progressive in their outlook than the liberal Canadian Mennonites.” Interestingly, Quiring (1928: 48) observes that the second wave of Mennonite migration to Canada had a comparable impact on Mennonite settlements in the Soviet Union, with vacated Mennonite land holdings opening to non-Mennonite settlement, thus eliminating the ‘closedness’ of the colonies.

tensions between these communities arising not only from individual reports of unfair dealings in exchanges of land between outgoing *Kanadier* and incoming *Russländer*, but also more fundamental differences in matters of education, accommodation, and integration into the surrounding society (Friesen 1988: 15–17). While both groups insisted on the importance of retaining their traditional languages, there is less evidence of systematic resistance to governmental intervention in Mennonite education in Canada on the part of *Russländer* groups. The belief that governmental involvement in education was incompatible with core aspects of the Mennonite social order was thus not shared in the same way by all denominations, and the resistance of *Reinländer* and *Bergthaler* Mennonites to government control over primary education was not generally reflected to the same degree among other Mennonite groups.

Under the new educational system to which *Kanadier* and *Russländer* groups were subjected, Mennonite children were required to attend English-language provincial schools, receiving at most one hour of education in High German per day (Ens 1994). The consequences of this arrangement were soon felt in the maintenance of religious and linguistic practices for Mennonite *Gemeinden*. The structure of this new education, no longer centred around the texts and linguistic practices with which adult members of the community were expected to be proficient, left central aspects of the traditional curriculum unattended, presenting a significant break in centuries-long educational traditions. In particular, the lack of substantive instruction in High German left younger members without the same facility in this language as their parents, with many unable to understand the content of church services, hymns, and even the standard catechism to the same degree as previous generations. While the affected communities eventually responded by organizing Sunday schools and private instruction in High German, this shift nevertheless eroded the traditional bilingualism that had been maintained since Mennonite immigration to Poland. Moreover, in many of the new provincially administered schools, children were explicitly forbidden from using *Plautdietsch* on school grounds, with the understanding that this would impede their acquisition of English (cf. Friesen 1988: 22). Marginalization of both traditional Mennonite languages in the educational realm contributed in the following years to a transition in Saskatchewan Mennonite churches to the use of *Plautdietsch* and, for all but the *Reinländer* Mennonites, eventually English for sermons and hymns (cf. Dawson 1936: 158–159, Doell 1987: 60).

Arriving in the wake of these changes, the Great Depression placed significant pressure on the agricultural basis of the Saskatchewan Valley settlements, as well. Continued concerns over educational freedom and the prospects for agricultural expansion prompted further migration on the part of Bergthaler and Old Colony groups into arable regions of northern Saskatchewan during the Great Depression, the Peace River area of northern Alberta in the mid-1930s, and, later, into Belize (1951) and Bolivia (1962). These migrations served to transplant Saskatchewan varieties of Mennonite Plautdietsch with each such move, often establishing communities in which Mennonite Plautdietsch remains to this day the language of daily interaction in essentially all domains (cf. Guenter et al. 1995: 419–421). The combined effect of depopulation from conservative Mennonite emigration, coupled with the economic pressures of the Depression, brought about the insolvency and final collapse of the traditional Mennonite social security system in the Saskatchewan Valley, placing Mennonite communities in closer contact with governmental agencies administered by an outside, anglophone majority and lessening overall community reliance on internal networks of mutual support centred in the *Gemeinden* (Doell 1987, Janzen 2004). Similarly, the severity of the Great Depression and the increasing mechanization of agriculture diminished the tenability of small-scale farming as a source of employment in many communities, providing incentive for younger Mennonites to enter the wage labour market in nearby urban centres. This process of urbanization represented a significant break from the traditional Mennonite economic system by which communities had formerly maintained their autonomy and cohesion, and presented a significant challenge for the maintenance of Mennonite linguistic and religious practices (Regehr 1996).

The advent of the Second World War further contributed to processes of linguistic and cultural shift that were already underway in these communities, bringing increased pressure on groups and individuals to align themselves with the British-Canadian mainstream to evade anti-German and anti-pacifist sentiments and prompting an intense debate in many Mennonite communities over non-participation in the war effort (cf. Doell 1987, Ens 1994). These internal divisions, combined with the aforementioned external pressures, had consequences for language use. Doell (1987: 60–61) reports that “numerous Mennonites stopped speaking Low German” during this period, with some individuals going so far as to changing their names to escape both harassment and a general sense of unease and even shame concerning their Mennonite roots.

All of these processes and pressures in concert—the break-down of traditional settlement patterns, the forced closure of Mennonite schools, the dissolution of Mennonite social welfare institutions, multiple waves of mass depopulation and repopulation, the opening of Mennonite reserve lands to non-Mennonite settlement, the effects of the Great Depression and agricultural mechanization on the economic basis of the community, urbanization, and the combined toll of two world wars—contributed to a shift from traditional multilingualism in High German and Plautdietsch towards monolingualism in English even before the onset of the Second World War. This is the situation reported by Brednich (1977: 9–10), who comments in his research on Mennonite folklore on rapid sociolinguistic change in these communities:

The main cultural change in the life of the Canadian-German Mennonites is linguistic. Only a few decades ago the Saskatchewan Valley used to be an exclusively German-speaking ‘island’. Low German was the everyday language and High German was spoken in church. German has become less and less important especially to the third generation growing up today. Through the influence of school children English has replaced German in an increasing number of families and today only some thirty percent of the children are able to understand Low German. [...] Generally speaking, one can say that the entire region is in a state of linguistic transition from German to English.

This parallels the situation noted in many other Russian Mennonite communities in Canada and the USA, where Plautdietsch-English bilingualism is almost universal and a shift to the latter language is seen in essentially all domains (cf. Buchheit 1982, 1988; Moelleken 1994).

Yet, this period was not exclusively one of language shift and loss. At the same time as Brednich’s report, significant anniversaries of Russian Mennonite settlement in western Canada and the emergence of national policies of multiculturalism brought renewed and largely positive attention to the historical and cultural distinctiveness of these communities. Although preceded by other notable Mennonite authors (e.g., Arnold Dyck; Suderman 1969; Reimer 1984, 1991), this period also saw the emergence of a significant body of written material in Mennonite Plautdietsch, with strong representation from Saskatchewan Valley writers such as Reuben Epp (e.g., Epp 1972, 1993, 1996), J. G. Janzen (cf. Guenter et al. 1995: 694–695), and, more recently, Jacob M. Fehr (e.g., Fehr 1993, 2001, 2006) and Jacob M. Driedger (Driedger 2011) whose works placed Plautdietsch in prestigious domains that had previously been the purview of High German (Reimer, Reimer & Thiessen 1983, Loewen & Reimer 1985, Friesen 1988: 22, Urry

1991). Through so-called *plautdietsche Owends* ‘Plautdietsch evenings’, full-length theatrical performances, sketches, poems, and songs were regularly performed in Plautdietsch in communities throughout western Canada, with several such events being organized by groups in the Saskatchewan Valley (Glendinning 2006). Likewise, weekly, half-hour broadcasts of Plautdietsch-language evangelical radio programmes (“Die Evangelische Botschaft”, later renamed “The Gospel Message”) have been produced by associates of the Evangelical Mennonite Mission Conference in Saskatoon since 1961 and are regularly broadcast throughout western Canada and internationally (Wiebe 2013). In short, while the linguistic shift observed by Brednich (1977) was no doubt significant, it was also not entirely uniform, as evidenced by these expansions of language use into domains of public performance, broadcast, and writing.

Although it is difficult to obtain an accurate assessment of the number of Plautdietsch speakers in the Saskatchewan Valley today, it is nevertheless possible to advance an estimate based on recent census data. It should be borne in mind that these figures may be inflated due to the presence of other speech communities in these districts who share the same linguistic classification in the census, or they may underestimate the size of the local Plautdietsch speech community by excluding individuals now living in nearby urban areas. Table 1 presents a summary of varieties in the Saskatchewan Valley classified as ‘German’ in the 2011 Canadian Census and the proportional representation of ‘German’ speakers in the overall population of the area. These same statistics are represented visually in Figure 3.

<i>Location</i>	<i>Pop.</i>	<i>L1 'German'</i>	<i>HL 'German'</i>	<i>Retention</i>
Aberdeen	600	10 (2%)	0 (0%)	0%
Aberdeen No. 373	1,010	55 (5%)	30 (3%)	55%
Blumenthal	65	15 (23%)	10 (15%)	67%
Corman Park No. 344	8,335	635 (8%)	375 (4%)	59%
Dalmeny	1,660	90 (5%)	10 (1%)	11%
Hague	875	175 (20%)	90 (10%)	51%
Hepburn	560	90 (16%)	40 (7%)	44%
Laird	285	50 (18%)	25 (9%)	50%
Laird No. 404	1,240	335 (27%)	250 (20%)	75%
Langham	1,255	70 (6%)	30 (2%)	43%
Martensville	7,720	205 (3%)	80 (1%)	39%
Neuanlage	140	45 (32%)	25 (18%)	56%
Neuhorst	125	30 (24%)	20 (16%)	67%
Osler	1,085	150 (14%)	105 (10%)	70%
Rosthern	1,560	190 (12%)	70 (4%)	37%
Rosthern No. 403	1,950	630 (32%)	480 (25%)	76%
Waldheim	1,000	140 (14%)	40 (4%)	29%
Warman	7,035	585 (8%)	315 (4%)	54%
<i>Total</i>	36,500	3,500 (8%)	1,995 (8%)	49% ( <i>avg.</i> )

*Table 1.* Population of Saskatchewan Valley communities and numbered rural municipalities (2011 Canadian Census). Estimates the total local population, the number of mother tongue speakers of varieties of German (L1), the number of speakers who use these varieties regularly in the home (HL), and a derived rate of overall language retention (HL / L1).



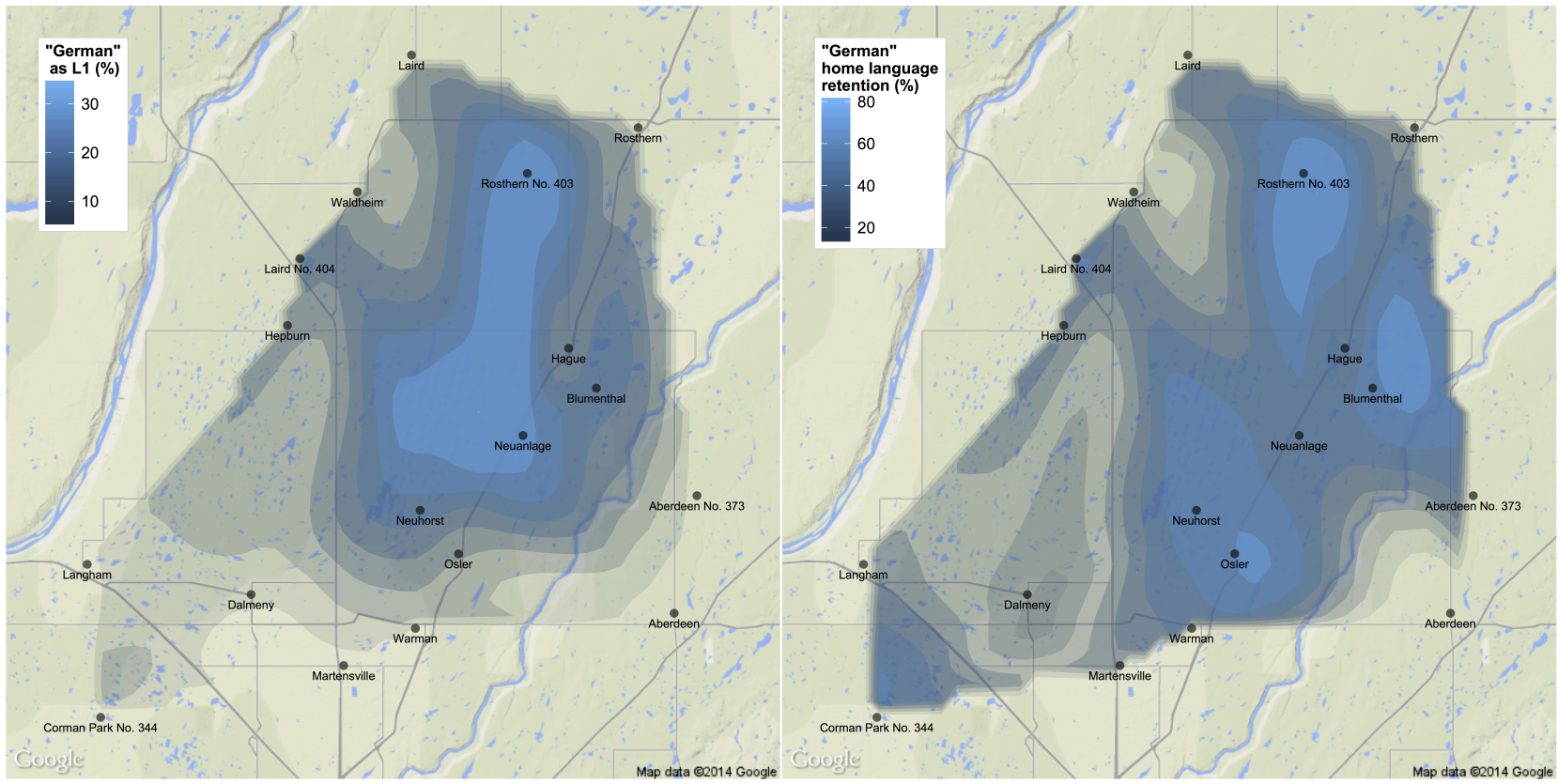


Figure 3. Proportion of first-language (L1) speakers of varieties of German in the Saskatchewan Valley (left panel), and rates of retention of German as a language of regular use in the home among these speakers (right panel; 2011 Canadian Census).

Several observations can be made on the basis of these data. Even under the assumption that all individuals who responded in the census that ‘German’ was among their languages of heritage or regular home use were referring to Plautdietsch—which would seem doubtful, given the presence of both nearby High German settlements and a High German-speaking Mennonite minority in some parts of the Saskatchewan Valley—this group is still in the minority, making up less than ten percent of the overall population of the area in 2011. Stronger proportional representation of speakers is found in the rural municipalities of the northern part of the Saskatchewan Valley, where between a quarter and a third of the rural population in the Rosthern and Laird areas report ‘German’ as their first language, as well as in some smaller towns and *Strassendorf* villages. A positive correlation is noted between the proportional representation of mother tongue speakers in a census area and the rate of language retention in the home ( $r_s(308) = 0.6821, p = 0.002415$ ). While adequate for such estimates, these census data are otherwise quite limited, providing no information on the distribution of speakers across age groups that might give a clearer sense of the progress of ongoing language shift.<sup>19</sup> Nevertheless, these data suggest that one or more of the traditional languages of local Mennonite communities are still spoken by an appreciable minority in the region, with the greatest vibrancy found in rural areas and smaller towns where a sizeable segment of the population shares them as their first language.

## 2.7 Summary

The history of the Dutch-Russian Mennonites is one of complex, multilateral migration and internal differentiation, punctuated by typically brief periods of stable settlement and inter-group contact. The Mennonite communities in the Saskatchewan Valley are no exception to this

---

19 By comparison, Anderson (2005: 21) estimates “over 10,000 people of German origin” in the Saskatchewan Valley in 2005. If accurate, this would suggest a considerably larger ethnic population in the area with no first-language experience of either Plautdietsch or High German, and thus give a sense of the magnitude of local language shift. Although the 2011 Canadian Census provides no information on ethnicity in its community profiles for direct comparison, it is interesting to note that Anderson finds “the population claiming German origin in [the Saskatchewan Valley] ranged from over 90% in Warman (92.4%), Hague (90.9%), and Osler (92.3%); over 80% in Martensville (83.8%), Waldheim (87.2%), and Hepburn (88.7%); over 70% in Dalmeny (77.6%) and Laird (77.1%); and a majority or close to it in Langham (58.7%), Rosthern (47.0%) and Aberdeen (41.8%)” in 1971.

pattern, though they present a particularly striking example of the manifold threads of relocation over several centuries coming together in a single place and point in time. With each successive relocation in this extensive migration history—whether from areas of northwestern Europe to northern Poland, or from separate communities in Poland to disparate colonies in Russia and Ukraine, or from these colonies to the USA, Canada, and beyond—and with each instance of contact or separation along other internal dividing lines, there exists the potential for variation to be introduced into the linguistic varieties maintained. The sudden divergence of previously united groups presents an opportunity, whether ultimately realized or not, for linguistic heterogeneity to develop in response to different internally and externally driven processes of language change. As several generations of scholarship have emphasized, the complexity of this history is thus evident in the substance of these communities' languages, presenting a palimpsest of these migrations and attendant processes of linguistic divergence and convergence over the course of a long, diasporic history.

The extent to which such processes are also reflected in linguistic developments in the Saskatchewan Valley, whether in the continuation of earlier dialect divisions or their levelling through contact and accommodation, and even the range of varieties present in this historically diverse settlement, is poorly understood. This is regrettable, not only in light of the historical importance of the Saskatchewan communities as a primary source of many thriving Plautdietsch speech communities in Latin America, but also in understanding more generally how the dynamics of language contact and change are reflected in these communities and their linguistic development. At the same time, the Saskatchewan Valley communities pose several challenges for traditional assumptions in dialectology. First, these speech communities are diasporic and colonial, and represent relatively recent settlements in western Canada. Thus, differences between communities may more closely reflect features of the linguistic situation of their previous points of origin, rather than the present-day geography in which the communities exist. Second, as is evident in the history of the Saskatchewan Valley communities, individuals and families are not geographically fixed, but have often moved to expand onto available lands elsewhere in the settlement area or left the settlement area entirely. As with the *Kanadier* emigration and *Russländer* immigration in the 1920s, the effect of such population movement can be a substantial shift in the demographic composition of large areas of the settlement region,

or less pronounced, as with the subsequent private sales of land between individuals of Mennonite and non-Mennonite backgrounds. Third, a long history of contact exists between Saskatchewan Valley Mennonite communities, extending from shared churches and schools in the earliest days of settlement to coincidental neighbours in district-based settlements and mixed villages. While this contact was limited historically by denominationally homogeneous settlement patterns, marriage between individuals of the same or similar *Gemeinde* background, and the maintenance of denominational private schools, the weakening of institutions maintaining this separation in later years, combined with the lessening of geographical constraints on interaction with improved transportation, might be expected to have consequences for patterns of interaction. This recent history of geographical, social, and denominational mobility that blurs the lines between historically distinct communities of practice in the Russian Mennonite diaspora, coupled with the reported exuberance of variation maintained in it, thus presents a series of theoretical and methodological problems for linguistic inquiry that are taken up in the following chapters.

### 3 Variation in Mennonite Plautdietsch

#### 3.1 Introduction

As the preceding chapter has detailed, the internal diversity encountered in contemporary diasporic Mennonite communities such as those in the Saskatchewan Valley reflect a complex history of contact and separation over several centuries of international migration. This diversity is reflected in the linguistic practices of these communities, as well, albeit without widespread scholarly consensus on either the diachronic sources or synchronic distribution of the linguistic features that differentiate constituent groups. How this history of interaction and isolation is manifested linguistically in these communities, and how this bears on the linguistic geography of the Saskatchewan Valley, in particular, presents open questions for linguistic research, and represents the focus of this chapter.

It is important to note at the outset that the historical and demographic factors discussed in this chapter as possible explanations for observed patterns of linguistic variation are often not entirely independent of one another. Rather, given the tight intertwining of ethnicity, religion, and migration in the history of the Russian Mennonites, it is more often the case that no clean separation can be made between settlement geography, migration history, socioeconomic status, and denominational affiliation, rendering comparisons of competing claims concerning these factors' relevance much more difficult.<sup>20</sup> Similar problems arise when comparing communities identified in different hypotheses, whether synchronically (e.g., when researchers adopt different criteria to distinguish speaker groups, thereby splitting the same population along different axes) or diachronically (e.g., when claims made concerning communities in one time period are compared against observations in the same communities at a later point in time). In such cases, some care must be taken to situate such claims in their specific historical and cultural framing, and not to extend them without warrant to situations other than the ones their proponents intended. In general, it should be understood that all such factors are at best imperfect proxies to recurring patterns of contact between individuals in particular linguistic communities of practice

---

<sup>20</sup> If denominational affiliation and settlement region are found to be highly correlated, for instance, it may be difficult to determine whether or not an analysis that proposes denominational differences to be of primary linguistic importance is substantively different from another that places more explanatory weight on geographical separation.

in particular social settings at particular points in time—and, as such, are neither necessarily mutually exclusive nor easily interpreted in isolation from one another.

Bearing this in mind, the following section aims to identify major families of hypotheses concerning the development and distribution of linguistic differentiation in Mennonite Plautdietsch, summarizing both proponents' arguments for their positions and offering critical assessment of their proposals. Drawing on the observations made here, the following chapter further considers the practical implications of these analyses and the methodological perspectives of Section 1.2 for the investigation of synchronic variation in the Saskatchewan Valley.

### *3.2 Hypotheses on Mennonite Plautdietsch differentiation*

Given the complex history that has contributed to the development of the linguistic repertoire of the Russian Mennonites, one might expect to find a range of perspectives in the scholarly literature on the events and processes implicated in the emergence of differentiation between varieties of Mennonite Plautdietsch. Before turning to consider areas of divergence among these views, this section begins by discussing two points of relative consensus: the acculturation of early Mennonite settlers in northern Poland to local vernacular norms (§3.2.1), and the existence of systematic linguistic differences between the two primary Mennonite colonies in Ukraine (§3.2.2). Following this, several hypotheses are reviewed which offer explanations—sometimes competing, sometimes complementary—for the maintenance or loss of such differentiation between Mennonite Plautdietsch varieties. These hypotheses include the geographical and social separation of Russian Mennonite colonies from one another (§3.2.2), denominational distinctions between Mennonite groups (§3.2.3), the passing of time between major periods of emigration and differences in settlers' socioeconomic status (§3.2.4), and different points of origin in the original northern Polish dialect landscape (§3.2.5).

#### *3.2.1 Early Mennonite adaptation to local vernacular norms*

Most proposals that have been advanced to account for linguistic differences between varieties of Mennonite Plautdietsch focus on time periods after the initial Mennonite relocation into northern Poland. While scholars differ in their treatments of the linguistic shift attributed to Mennonite refugees entering northern Poland in the sixteenth century, there is general agreement

that the transition to vernacular use of local varieties of Plautdietsch left few traces of the earlier linguistic diversity of these newcomers (except in a limited number of possible substratal contact features; cf. §3.2.5). Given the scarcity of vernacular linguistic records from this period, it is difficult to entertain more than speculation on the processes of linguistic shift and accommodation that may have taken place among Mennonites in the region to limit the preservation of earlier linguistic distinctions in later varieties of Plautdietsch. Nevertheless, it is commonly noted that elements of Dutch and Flemish (and possibly Frisian, as well, although this remains contested; see Nieuweboer & de Graaf 1994 and Siemens 2012: 62–68 for discussion of several possible lexical and grammatical influences) came to be incorporated into varieties of Plautdietsch throughout the region. Rather than being characteristic of Mennonites in the Vistula region alone, however, such northwestern European influences appear to have been distributed throughout the general population (cf. Wiens 1916, Ziesemer 1924). Mitzka (1930: 12) observes that Mennonite speakers of Plautdietsch acquired “the respective local form of the dialect” (“die jeweilige örtliche Form der Mundart”) in northern Poland, such that there was no “specifically Mennonite variety of Plautdietsch extending over the entire Vistula region” (“ein besonderes, etwa über die ganze Weichsellandschaft gelagertes Mennonitenplatt”).

The apparent receptiveness of Mennonite settlers in northern Poland to local vernacular norms is evidenced in the maintenance of these distinctive varieties in descendant Mennonite communities, as well. In the earliest report on a Russian Mennonite speech community, Baumann (1854: 441) observes significant dialect differences in the Molochnaya Colony in the mid-nineteenth century, to the extent that occasional difficulties in mutual comprehension are noted. These divisions are confirmed to exist some seventy years later by Quiring (1928: 44–45) (albeit with one community having shifted to the use of another variety), and persisting even in daughter settlements in Kansas well into the twentieth century (Krahn 1959: 187). Closer inspection of the five varieties identified by Baumann suggests clear connections to the settlement histories of their associated *Gemeinden*, with several of these groups having settled outside of the core West Prussian dialect area in which most other Mennonites were located, thus maintaining their coherence through later migrations en masse (cf. Siemens 2012: 47). In all such cases, Mennonite groups appear to have adopted the local vernacular conventions of the region of their initial settlement, whether in northern Poland or elsewhere, and subsequently

maintained these practices even across significant migrations. Examples such as these provide a sense of the degree of initial receptiveness of Mennonite migrants to the adoption of local speech norms in northern Poland, as well as the apparent level of entrenchment of these varieties among individual groups in later settlements.<sup>21</sup>

### 3.2.2 *Differences between Chortitza and Molochnaya Colonies*

Whether such distinctions were ultimately maintained or abandoned, it is generally agreed that Mennonites arriving in Ukraine brought with them a number of Polish Plautdietsch varieties, and that this diversity contributed to differences in the forms of speech later associated with the Chortitza and Molochnaya Colonies (cf. Epp 1993: 77). Nevertheless, there is somewhat less consensus on the role that the physical separation of these two colonies from one another may have played in the emergence of distinctive vernacular norms. Thiessen (1989: 285, fn. 1) cites several researchers (e.g., Quiring 1928, Lehn 1957, Dyck 1964, Mierau 1964, Goerzen 1972, i.a.) as supporting “to a greater or lesser degree” the position that differences between colonial Mennonite Plautdietsch varieties are the result of the separation of the two colonies.

The hypothesis that the geographical division of the two primary Mennonite settlements in Ukraine was itself a contributing factor in the development of later dialect diversity is prominent in the literature on Mennonite Plautdietsch and appears in several forms. In the stronger version of this colonial separation hypothesis, linguistic differences between the Chortitza and Molochnaya Colonies are not attributable to an earlier state of dialect diversity to which Mennonite immigrants would have been exposed, but rather to independent developments in each of the colonies in Imperial Russia. This is reportedly the position of Viktor Schirmunski (noted in Thiessen 1989: 295–296), who suggests that such dialect differences were primarily the product of linguistic divergence between the two colonies after Mennonites’ immigration to Ukraine. This strong hypothesis is different from the more moderate version advocated by the

---

21 Outside of the Molochnaya and Chortitza Colonies, one also finds Polish Mennonites who emigrated from the more southerly Kulmerland region near present-day Grudziądz (Graudenz) and Chełmno (Culm) to the region of Volhynia east of Kiev in Ukraine (Epp 1993: 81–82). Paralleling other groups, these Mennonites reportedly adopted the local variety of Eastern Pomeranian (Kulmerland) in Poland, which they later maintained in Volhynia and in subsequent migrations to Kansas and South Dakota.



other scholars cited in Thiessen (1989). On their view, the separation of these colonies may well have contributed to linguistic divergence between them over time, but such differences did not proceed from a state of linguistic uniformity at the time of initial settlement. The early reports of Baumann (1854) and Quiring (1928) on Mennonite settlements in the Russian Empire favour this position, bringing attention to significant dialect divisions that cannot be easily attributed to independent linguistic developments in the immediate post-immigration period. Without ruling out the possibility of later, independent linguistic developments in both colonies, and in the absence of a distinctively ‘Mennonite’ variety of Plautdietsch at the time of Mennonite immigration to Ukraine, it is arguably more plausible to assume the weak hypothesis that elements of the dialect continuum in northern Poland were transplanted into these new colonies, whatever their ultimate fate.

From a sociolinguistic perspective, the separation of Russian Mennonite settlers into two primary colonies appears to have had several immediate consequences. On the one hand, the geographical separation of one colony from the other presented an opportunity for linguistic divergence to develop between their respective vernacular norms, as argued by later scholarship that treats the Chortitza-Molochnaya division as a linguistically significant one (cf. Mitzka 1930: 13, 21; Nyman 1997: 266).<sup>22</sup> On the other hand, the relative isolation of these colonies appears to have also encouraged greater interaction between the settlers within them, and thus also encouraged the levelling of linguistic distinctions between groups within each colony, such that linguistic minorities in the Mennonite population gradually assumed the vernacular norms of demographically more dominant groups (where confessional distinctions or other factors did not

---

22 While Nyman (1997: 266–268) divides the varieties of Mennonite Plautdietsch in the Orenburg settlement in western Siberia into Chortitza and Molochnaya varieties, he nevertheless notes that “this strict, ideal Chortitza-Molochnaya dichotomy does not entirely correspond to linguistic reality” (“[d]iese strenge, ideale Dichotomie Chortitzaisch/Molotschnaisch entspricht aber nicht ganz der sprachlichen Realität”; Nyman 1997: 268). In his view, while the Molochnaya variety can be considered a relatively homogeneous dialect, the Chortitza variety “appears in multiple distinct gradations which are temporally, spatially, and possibly also confessionally established” (“erscheint in mehreren unterschiedlichen Abstufungen, die zeitlich, räumlich und vielleicht auch konfessionell begründet sind”). Interestingly, Nyman’s observation of greater dialect diversity in the Chortitza group is the opposite of what has often been claimed by other scholars (e.g., by Quiring 1928: 44–46).

discourage this; see below).<sup>23</sup> The centralization and shared administrative (and, later, also health and educational) systems in the colonies, coupled with close economic and personal contact between individual families in each settlement (cf. Quiring 1928: 44), presented conditions favourable to language shift. Whereas the geographical separation of colonies may have contributed to linguistic divergence between them, there is relatively clear historical evidence of widespread linguistic convergence within each colony, resulting in an overall reduction in dialect diversity over successive generations of settlement in Ukraine (cf. Quiring 1928: 45, fn. 63).

### 3.2.3 *Differences between denominations*

In addition to the separation of the Chortitza and Molochnaya colonies, Quiring (1928: 42ff.) proposes denominational differences as a linguistically relevant dividing line between speakers of Mennonite Plautdietsch. Quiring relates much of his discussion of dialect differences in both colonies not only to their respective migration histories, for which he provides extensive documentation, but also to the long-standing confessional differences between ‘Frisian’, ‘Flemish’, and ‘Old Flemish’ Mennonites described in Section 2.2. Quiring (1928: 43) is explicit in his hypothesis that the proclivity of each of these denominations to disfavour interaction with one another lessened the potential for dialect mixture and levelling, ultimately preserving existing linguistic differences between these groups. Thus, Quiring is careful to note that the two primary dialects he observes in the Chortitza Colony correspond to earlier confessional differences between their speakers. According to him, the “Chortitza” dialect is spoken by the descendants of Flemish Mennonites, and the “Frisian” dialect by the descendants of a separate group Frisian Mennonites in the villages of Kronsweide, Schönwiese, Kronsgarten, and Einlage. With the reasons for the historical differences between these groups having since been forgotten and opposition between both factions essentially non-existent, Quiring (1928) claims, both groups subsequently entertained closer contact with one another, and the minority Frisian varieties consequently began to give way to the majority Chortitza dialect.<sup>24</sup>

---

23 The terms ‘confessional’ and ‘denominational’ are used interchangeably in this study.

24 Although he does not relate this division to confessional differences specifically, Schirmunski (1928: 53ff., cited in Thiessen 1989: 296) also observes a distinction between the majority dialects of the Chortitza and

The linguistic relevance which Quiring (1928) attributes to denominational divisions has been repeatedly called into question, both by his contemporaries and by later scholars. The earliest critique of this position is offered by Mitzka (1930: 22–23), who argues that confessional differences between Flemish and Frisian Mennonite groups cannot be considered explanatory of linguistic differences among the Russian Mennonites. Mitzka points out that both the Chortitza and Molochnaya colonies were founded by Flemish Mennonites, with Frisian groups representing either small minorities among early settlers or a larger proportion of latecomers, and often living in mixed villages with non-Frisians. Despite the shared denominational background of the primary founders of both colonies, Mitzka argues, dialect variation between both colonies nevertheless exists, suggesting that explanations must be sought somewhere other than in confession. Moelleken (1987: 93) also finds little evidence in favour of Quiring’s conclusion, noting that confessional differences failed to effect a distinction between Mennonite and non-Mennonite varieties of Plautdietsch in northern Poland, let alone between elements of the Polish Mennonite population itself. Both Thiessen (1989) and Siemens (2012: 47–48) arrive at similar conclusions, with Siemens suggesting that the linguistic evidence furnished by Quiring to support a ‘Frisian’ vs. majority Molochnaya division is not conclusive, and that the separation between these communities was likely confessionally, rather than linguistically determined.

Even while acknowledging that the evidence for clear-cut linguistic differences extending primarily along confessional lines in the earliest Mennonite settlement period in Ukraine is at best scant, leaving the linguistic situation open to other interpretations, this does not preclude denominational affiliation from proving linguistically relevant in other cases, either historically or contemporarily. With the linguistic differences between Mennonite groups in the Molochnaya Colony noted by Baumann (1854) persisting over several generations and even into settlement in North America, one might speculate that the successful maintenance of these distinctive varieties was supported in part by the presence of distinct *Gemeinden* in each of these settlements. At the very least, there is not sufficient evidence to preclude denomination as a potential contributing factor in the maintenance of distinctive local speech norms in this period of Russian Mennonite history. Nor is there reason to assume that such differences are necessarily irrelevant in

---

Molochnaya colonies and what he calls the “Kronsweide dialect,” lending some independent support to Quiring’s differentiation of the Chortitza Frisian group to which Kronsweide belongs.

descendant communities. Both Dyck (1964) and Brandt (1992) identify linguistic differences between particular Russian Mennonite denominations in Manitoba and Mexico, respectively, although several of these distinctions could also be attributed in part to correlated differences in migration history. In both these historical and contemporary cases, the linguistic relevance of denominational differences in distinguishing varieties of Mennonite Plautdietsch cannot be dismissed out of hand on the basis of sparse evidence in one historical context, however prominent that instance may be.

### 3.2.4 Differences in emigration period and socioeconomic class

While denominationally influenced developments in Mennonite Plautdietsch may be traced back to particular historical circumstances with relatively little difficulty, accounting for the dialect divisions noted in the Ukrainian colonies clearly requires attention to other factors. For a start, the earliest literature on Mennonite Plautdietsch dialect variation repeatedly notes that the variety of Plautdietsch later associated with the Chortitza Colony is not attested anywhere in the northern Polish dialect landscape from which it emerged (Quiring 1928, Mitzka 1930). While some scholars attribute this discrepancy to processes of dialect contact and mixture in Imperial Russia, which presumably contributed to the emergence of novel configurations of earlier geographically associated linguistic features (cf. Quiring 1928: 42), this position is not uncontested. Mitzka (1930) is openly critical of this ‘mixed dialect’ hypothesis, and instead connects such differences to posited historical changes in the northern Polish dialect landscape that affected the distribution of varieties that Mennonite settlers brought with them to Ukraine. Under Mitzka’s hypothesis, the ascendance of a Low German trade dialect during the period of Mennonite emigration from Poland radically altered the dialect geography of the region, leaving the more isolated eastern *Nehrung* as a linguistic ‘relict area’ (*Reliktlandschaft*) where the influence of this variety was less pronounced (Mitzka 1930: 13, 21; Mitzka 1968b [1924]: 214–215). Citing homologies between Chortitza Colony varieties and varieties in this region on the one hand, and between Molochnaya Colony varieties and varieties of Plautdietsch spoken elsewhere in northern Poland on the other, Mitzka concludes that the earlier Chortitza settlers brought with them to Ukraine varieties of Plautdietsch that were little affected by the ascendant trade dialect, and, conversely, that Molochnaya settlers brought with them dialectally shifted

forms in later years.

Importantly, Mitzka accounts for these differences not only by reference to the temporal division between the establishment of both Russian Mennonite mother colonies and concomitant differences in exposure to posited regional processes of language change in northern Poland, but further asserts that socioeconomic factors were of central importance to later linguistic differences (Mitzka 1930: 13, 23). Mitzka founds this hypothesis on the claim that the first settlers arriving in Chortitza were on the whole poor and not well educated, and thus would have had limited contact with either the dominant written language (High German) or the language of commerce in the Vistula region. In contrast, Molochnaya colonists emigrating from northern Poland after 1803 were assertedly of higher social standing and greater wealth, and thus brought with them linguistic forms that had been significantly influenced by both commerce and education, due as much to their longer stay in the Vistula region during a period of major linguistic change as to their contact with trade dialect speakers. Consequently, Mitzka sees little need to introduce dialect mixture or settler denomination as further explanatory factors. Seen through this lens, the linguistic situation admits an explanation in which Chortitza varieties are equated with the “dialect of the northern Great *Werder* and its northern neighbours, maintained in a remote location, as it was represented by the lower classes of speakers at the end of the 18th century” (“in der Ferne festgehaltene Mundart der Gegend des nördlichen Großen Werdens und seiner nördlichen Nachbarschaft, wie sie Ende des 18. Jahrhunderts durch die untere Schicht der Sprachträger vertreten war”; Mitzka 1930: 21), and Molochnaya varieties to later, higher class and more trade language-influenced forms of Plautdietsch.

This explanation is compelling, particularly in light of Mitzka’s extensive scholarship on the dialect geography of the Vistula region (e.g., Mitzka 1968a [1922], 1968b [1924]), and has been of considerable influence on later linguistic studies of Mennonite communities (Lehn 1957, Dyck 1964: 64–65, Tolksdorf 1985: 327–328, Thiessen 1989: 286, 288; cf. Moelleken 1987: 98). However, more recent historical scholarship has questioned the extent of the socioeconomic and educational divisions that are commonly assumed to have separated the earliest Chortitza and Molochnaya settlers and for which Mitzka (1930) provides no evidence. Both Plett (2000, 2001) and Schapansky (2006) raise significant doubts as to the factuality of pronounced economic differences between early Chortitza and Molochnaya settlers, providing reanalyses of available

historical economic data to suggest considerably less socioeconomic inequality existed between both groups than was often suggested in earlier Mennonite historiography (cf. Epp 1974, Thiessen 1989: 291, fn. 12). This leaves the socioeconomic element of Mitzka's proposal in question: if such differences between early Russian Mennonite settlers are not as stark as previously thought, then it is unlikely that their respective degrees of exposure to a regional trade dialect or dominant written language would prove explanatory for such significant linguistic differences as are attested between both groups.<sup>25</sup>

Even if socioeconomic differentiation is not a sure assumption for either the earliest settlers in the Chortitza and Molochnaya Colonies (Schapansky 2006) or for those Mennonites who migrated to Canada in the 1870s (Plett 2000, 2001), this still does not preclude dialect shifts from having taken place throughout the Vistula region at the time of Mennonite emigration, and later dialect differences between Russian Mennonite colonies thus reflecting settlers' respective periods of emigration. However, as several scholars have noted, the assumption that widespread dialect shift is reflected in these linguistic differences is potentially difficult to square with both the narrow windows of time and the magnitude of the linguistic changes under consideration, with only fifteen years falling between the establishment of the primary Russian Mennonite colonies (Tolksdorf 1985: 327; cf. Epp 1993: 79). Moreover, Siemens (2012: 50) notes that Mitzka "provides no evidence whatsoever that the language in the Great *Werder* was developing particularly rapidly" in the period between Mennonite migrations ("er führt keinerlei Evidenzen dafür an, dass sich die Sprache im Großen Werders um 1800 besonders rasant entwickelte"). It is certainly possible that the linguistic landscape of the Vistula region may have been radically reconfigured in this brief period, whether in the wake of an ascendant trade dialect or for other, less well understood reasons, and that this in turn affected the vernacular norms of the Mennonite population in Poland. But the extent of such differences, reportedly occurring at all levels of

---

25 Moreover, acceptance of socioeconomic differences between early Chortitza and Molochnaya settlers does not immediately imply that these differences are linguistically relevant. Siemens (2012: 50) questions Mitzka's argumentation in this respect, pointing out that "even though a divide in affluence existed between the two groups, there are nevertheless no indications that different social classes in the same region in West Prussia spoke different forms of Low German" ("da zwar ein Wohlstandsgefälle zwischen den beiden Gruppen bestand, es aber keine Indizien dafür gibt, dass in Westpreußen die verschiedenen sozialen Schichten derselben Region unterschiedliches Platt sprachen").

linguistic organization, requires the assumption of a shift so rapid, so widespread, and so linguistically significant as to have affected disparate Mennonite populations in northern Poland equally in less than a generation's time, leaving whatever previous varieties these Mennonites may have maintained as linguistically inconsequential minorities (cf. Thiessen 1989: 286). In the absence of strong, corroborating evidence, such an assumption is difficult to consider plausible.

At the very least, such a marked and rapid transition as Mitzka (1930) argues had occurred during this period might be expected to be evidenced in the resulting linguistic landscape. As noted above, the primary evidence offered in favour of this posited historical shift is several correspondences between features of the majority Chortitza variety and of eastern *Nehrung* varieties on the one hand, and between the majority Molochnaya variety and much of the remaining Vistula region on the other. Combined with the seemingly paradoxical observation that the eastern *Nehrung* was the site of only minimal Mennonite settlement, these observations have commonly been interpreted to suggest that a linguistic shift must have taken place between waves of Mennonite emigration, relegating Chortitza-like features to the geographical periphery of the region and preventing them from being imported with most later Molochnaya settlers. Yet, several scholars have expressed reservations about this characterization of the local dialect landscape, noting broad, cross-regional variation even in supposedly 'characteristic' features (cf. Tolksdorf 1985: 327). Moelleken (1987) re-examines the dialect-geographical evidence offered by Mitzka (1930) for an exclusive correlation between features of the majority Chortitza variety and varieties in the eastern *Nehrung*, comparing Mitzka's description against historical data from the *Deutscher Sprachatlas* and synchronic documentation of descendant Polish and diasporic Plautdietsch varieties. Significantly, Moelleken (1987) finds both historical and contemporary attestation throughout the Vistula region of multiple linguistic features claimed by Mitzka (1930) to occur only in the eastern *Nehrung*, thus contradicting several of Mitzka's distributional claims. Apparent discrepancies between colonial Mennonite Plautdietsch varieties and the northern Polish dialect landscape that motivated Mitzka's hypothesized shift, Moelleken suggests, are most likely due to an "uneven and not representative data set" ("eine uneinheitliche und nicht-repräsentative Datenbasis"; Moelleken 1987: 89) ultimately contributing to dialect-geographical generalizations that simply "do not conform to the linguistic facts" ("entsprechen [...] nicht den

sprachlichen Gegebenheiten”).

These observations clearly do not remove the possibility of significant historical shifts in the Polish Plautdietsch dialect landscape, whether motivated by socioeconomic or other factors. They do, however, call into question the assumption that a fundamental reconfiguration of the northern Polish dialect landscape in the period between Mennonite migrations to the Russian Empire is prerequisite to any coherent account of variation between descendant speech communities. By comparison with Mennonite emigration from northern Poland, much more consensus exists concerning the linguistic relevance of differences in emigration period between Russian Mennonite groups in the Americas, which are typically seen as “valid and essential parameters” of variation (Epp 1993: 95). Several authors note greater influence from both Slavic languages and High German among *Russländer* Mennonite immigrants than among the earlier *Kanadier* Mennonites, as well as the increasing predominance and prestige of Molochnaya-associated forms across all varieties of *Russländer* Plautdietsch (Dyck 1964, Epp 1993; cf. Quiring 1928, Mitzka 1930).<sup>26</sup>

### 3.2.5 Differences in dialect geography and points of origin

Several hypotheses concerning variation in Mennonite Plautdietsch relate later linguistic differences to the disparate nature of earlier Mennonite settlement in the dialect landscape of northern Poland (cf. §3.2.1). While there is general agreement that geographically associated patterns of variation in this region likely exerted some influence on forms of speech encountered in descendant communities, it is less clear to what extent such variation can be directly attributed to particular historical settlements. Processes of divergence and convergence among descendant groups, coupled with recurring instances of ‘mixed’ settlement, may obscure or even eliminate direct dialect-geographical correspondences between features in historical and contemporary Plautdietsch varieties.

Such processes notwithstanding, several scholars have proposed that contemporary variation in Mennonite Plautdietsch may have more direct roots in Polish dialect geography than

---

26 On differences in language attitudes between *Kanadier* and *Russländer* Mennonites, see also Epp (1993: 90–91), who observes that assessments of the overt prestige of particular varieties of Mennonite Plautdietsch differed considerably between earlier and later Mennonite immigrants to Canada and the USA.



is sometimes assumed. It has been noted that the Mennonite population in northern Poland was situated predominantly in the Malbork *Werders*, with the overwhelming majority of immigrants to both the Chortitza and Molochnaya Colonies prior to the Napoleonic Wars being of the ‘traditionalist’ Flemish contingent from the *Nehrung* region or the northern half of the Great *Werder*, while later waves of migrants to the Molochnaya Colony included considerably more Frisian Mennonites from the central and southern Malbork and Elbląg *Werder* areas (Quiring 1928: 13, 16; Mitzka 1930: 7; Tolksdorf 1985: 328; Schapansky 2006: 143–145). Drawing on observations such as these and more detailed comparisons of specific linguistic features, both Moelleken (1987: 122) and Siemens (2012: 47ff.) propose a direct correlation between Chortitza Plautdietsch and the varieties of Polish Plautdietsch spoken in the *Nehrung*, and between Molochnaya Plautdietsch and varieties in the Great *Werder*.

Proposals of significant correlations between earlier settlement patterns, dialect geography, and the later distributions of dialect features in Mennonite Plautdietsch speech communities are not restricted to the relationship between northern Poland and the first Mennonite colonies in Ukraine. Siemens (2012: 62ff.) proposes that Flemish Mennonites entering northern Poland transferred several characteristics of their earlier varieties onto corresponding features of local Plautdietsch, leaving substratal traces in descendant varieties. Siemens cites homologies in East Flemish and Chortitza Plautdietsch vowel inventories, the common retention of *-en* infinitival endings and, perhaps most prominently, the typologically unusual rounding of West Germanic *\*au* to [yə] in both East Flemish and Chortitza Plautdietsch as evidence for such influence, arguing that these correspondences as a whole are suggestive of historical processes other than coincidental polygenesis.<sup>27</sup> Given the historical connection between the Dutch-Flemish and Frisian linguistic areas and the Mennonite populations entering northern Poland in the mid-sixteenth century, as well as the largely distinct regions in which the Flemish and Frisian Mennonite populations came to be centred (and the heavy commercial connections throughout the northern delta region to Dutch-speaking merchants), it is not unrealistic to propose that these linguistic histories might be reflected in substratal characteristics

---

<sup>27</sup> Moelleken (1993) arrives at a similar conclusion, albeit for a single feature, suggesting that the alveolar approximant (‘retroflex’) approximant allophone of /r/ found in some Chortitza-descended Mennonite Plautdietsch varieties may be of Dutch-Flemish origin.

of descendant Mennonite Plautdietsch varieties.

### 3.2.6 *Summary*

The historical factors that have been proposed to have contributed to the emergence of Mennonite Plautdietsch in the forms encountered today in areas like the Saskatchewan Valley are both complex and contested, both in the linguistic relevance of individual features and, occasionally, also the historiographical accuracy of the accounts on which such proposals are founded. As previous scholarship has sought to underscore, the interaction of multiple factors in the history of the Russian Mennonites—repeated, multilateral migration involving linguistically distinct settlements and regions; internal denominational, geographical, and socioeconomic divisions; and differing patterns of contact with (and attitudes towards) other linguistic communities with which Mennonites interacted—are all potentially relevant to the present state of internal linguistic diversity. Despite differences between their specific proposals, these same researchers have argued with essential unanimity that the historical and cultural frame in which linguistic developments took place has explanatory value beyond providing contextualizing detail. Hypotheses concerning the genesis of Mennonite Plautdietsch in its present form in descendant communities are intimately tied to the specific linguistic history of the Russian Mennonites as an internally diverse religious minority whose repeated relocation has placed them in no less diverse external linguistic settings.

Such influences, coupled with an extensive history of separation and contact between constituent speech communities, have had complex effects on the diachronic development of Mennonite Plautdietsch in the diaspora. Nevertheless, it is possible to arrive at a clearer picture of the overall linguistic situation when the hypotheses proposed to explain its emergence are assembled:

1. *Migration history.* While not independent of other factors, migration history is relevant to the linguistic characteristics of present-day Mennonite Plautdietsch in several respects. Differences in place of origin and period of emigration have been argued to have been linguistically important at essentially every point of Mennonite migration:
  - In the earliest period of Mennonite relocation into northern Poland, Mennonite newcomers came to adopt local varieties of Plautdietsch and generally maintained

- these across subsequent migrations (§3.2.1). The earlier linguistic practices of Flemish and Frisian Mennonites have been argued to have contributed several substratal features reflected in descendant varieties of Mennonite Plautdietsch (§3.2.5), although further correlations between these languages and specific varieties of Mennonite Plautdietsch have proven difficult to demonstrate conclusively, given the considerable time depth and extremely limited documentation of relevant vernacular features in historical records for the Vistula region.
- In later Mennonite movements into Imperial Russia, participation in particular waves of migration was relevant to the constitution of linguistic majorities and minorities in the major Russian Mennonite colonies (§3.2.4). The earliest Mennonite settlers in Ukraine were generally Flemish Mennonites from the northern delta (*Nehrung*) region, while later groups had more predominant representation of Frisian Mennonites from more southerly (*Werder*) areas. For some scholars, the time gap between the largest waves of emigration that established these first Russian Mennonite colonies is also seen as linguistically relevant, as a dramatic reconfiguration of the northern Polish dialect landscape, proposed to have taken place between these migrations, presumably affected the set of geographically associated linguistic features that Mennonite settlers later brought with them to Ukraine. Other research has called this proposal into question, noting only a narrow window of time between waves of settlement (typically less than a generation) during which such radical changes could have taken root among Mennonite emigrants. This research instead points to considerable regiolectal variation in Polish Plautdietsch and differences in the areas of that dialect landscape from which most Mennonites emigrated to explain later linguistic differences between Russian Mennonite colonies.
  - Finally, the separation of Russian Mennonite groups through migration to the Americas beginning in 1873 is repeatedly claimed to be linguistically relevant in later settlements. A sharp linguistic division is reportedly found in western Canada between earlier *Kanadier* and later *Russländer* Mennonites of both Chortitza and Molochnaya heritage (§3.2.4).
2. *Denominational divisions.* Although far from universally accepted, it is possible that

confessional differences between groups of Plautdietsch speakers may have limited their interaction with one another, and thus contributed to the maintenance of distinctive vernacular speech norms in some communities (§3.2.3). On this view, as differences between Mennonite denominations became less pronounced over time, greater interaction between groups and joint participation in shared social institutions may have contributed to increased convergence between their respective varieties of Plautdietsch.

3. *Socioeconomic differences.* Similarly, socioeconomic differences are sometimes contended to have affected the distribution of linguistic features in Mennonite settlements in Poland and Ukraine (§3.2.4). Although earlier scholarship posited considerable differences in relative affluence and education between earlier Chortitza and later Molochnaya Mennonites that were suggested to be linguistically relevant, more recent historical research calls this into question, finding less pronounced socioeconomic differences between successive waves of settlement than were once assumed. While social class may have correlated with linguistic divisions in both Poland and Ukraine, this has not been conclusively demonstrated, and there are no reports of deep-running internal linguistic divisions among 1870s Mennonite emigrants along socioeconomic lines. By comparison, increased interaction between Mennonite colonies and social institutions, changes in educational practices, and on the whole greater socioeconomic differentiation in the Mennonite settlements in Ukraine after 1873 reportedly had a significant effect on vernacular linguistic norms. These social changes may have contributed to the prestige of linguistic forms associated with the demographically and economically more prominent Molochnaya Colony, as well as greater influence from both High German and local Slavic languages (Russian, Ukrainian) on those Mennonite Plautdietsch varieties that would later be associated with 1920s *Russländer* emigrants to Canada.

As this synopsis makes apparent, there is no single line along which Mennonite Plautdietsch varieties can be cleanly differentiated. Several distinct axes of affiliation are potentially relevant to patterns of linguistic differentiation in descendant speech communities such as those in the Saskatchewan Valley, where earlier ethnic-religious affiliations, settlement and migration histories, and possibly also later denominational groupings may be of linguistic significance. Importantly, these axes are not independent of one another in Russian Mennonite history, with

prominent correlations between particular ethnic-religious groups and migration and settlement histories (e.g., Flemish-*Nehrung*-Chortitza, Frisian-*Werder*-Molochnaya; cf. Siemens 2012: 65). From an historical perspective, then, the present-day linguistic situation can be viewed as partly the result of the diachronic interaction of features such as these, none of which is entirely explanatory in isolation or completely orthogonal to all other predictors. Given the complexity of the individual histories of constituent communities, families, and speakers, it is clear that the distribution of distinctive linguistic features in contemporary Mennonite Plautdietsch speech communities can be explained only in part by such histories, which serve to varying degrees of effectiveness as proxies for larger patterns of interaction, convergence, and divergence in the speech community as a whole.

Contemporary Mennonite Plautdietsch speech communities such as those in the Saskatchewan Valley are thus difficult to characterize adequately with a single label, given the range of historical factors which are of potential relevance to the present linguistic situation. In most contemporary speech communities, it remains to be determined what the actual linguistic substance of extant divisions is; what, if any, correlation these divisions have with the aforementioned historical and demographic factors; and how feasible it is to arrive at a set of finer-grained distinctions between contemporary varieties of Mennonite Plautdietsch, especially in light of well-attested processes of linguistic convergence in speaker communities (cf. Nyman 1997). The following chapter explores these problems in more detail, considering how these issues might be approached methodologically to determine which, if any, of the proposed explanations for variation in Mennonite Plautdietsch prove relevant to the linguistic situation of the Saskatchewan Valley today.

## 4 Methodology

### 4.1 Introduction

The historical circumstances that led to the presence of a significant Russian Mennonite population in the Saskatchewan Valley today are complex, representing one consequence of the centuries of repeated migration, internal division, and contact that have culminated in the present global diaspora of Mennonite Plautdietsch speech communities. Even among these diverse Russian Mennonite settlements, the Saskatchewan Valley stands out as unusually heterogeneous and of particular comparative linguistic importance—both as a meeting place of several historically distinct streams of Mennonite migration in a single area, and as a point of departure for and one significant source of vibrant Mennonite Plautdietsch speech communities in northern Alberta, British Columbia, Kansas, Texas, and Latin America. As the preceding chapter has emphasized, the same historical circumstances have not been without consequence for the language of these settlers, as well. Several generations of linguistic scholarship involving diasporic Russian Mennonite communities attest to the complex interplay of settlement patterns, migration history, denominational affiliation, and other such factors in linguistic variation within and between Mennonite Plautdietsch speech communities. While such is the case elsewhere in the Russian Mennonite diaspora, it remains to be seen to what degree similar patterns of linguistic differentiation are also encountered in the particularly impressive internal diversity of the Saskatchewan Valley. When viewed in this light, the linguistic situation of these communities presents a series of interrelated problems for research:

1. Given the absence of any prior linguistic documentation in these communities, what is the extent of variation in present-day Saskatchewan Valley Mennonite Plautdietsch? Moreover, how might this variation be adequately represented, such that the results of this investigation both provide insight into the present linguistic situation and inform further documentary efforts?
2. Given a repeated insistence in the scholarly literature on the existence of significant associations between historical and demographic features of Mennonite communities and their respective forms of Plautdietsch, what correlations, if any, does linguistic variation in the Saskatchewan Valley enter into with the demographic and historical characteristics of these communities and their speakers?

3. Given an unknown number of varieties present in the Saskatchewan Valley, is it possible to identify coherent, recurring patterns of linguistic variation—clusters of conventional variant selection that may suggest distinct varieties? Moreover, as the relative importance of particular variables in defining the boundaries between speech communities in the Saskatchewan Valley is also unknown, is it possible to arrive at such divisions inductively—from observations of variation across over a number of predictors, none of which are favoured *a priori* as the basis of classification, rather than imposing them top-down through the selection of a limited set of features presumed to be of linguistic interest?
4. Given the larger societal context of language shift and loss in which the answers to these questions must ultimately be sought, how might linguistic research undertaken in these circumstances be consonant with the interests of local communities, producing outcomes that are of benefit not only to linguistic analysis, but also to local language education, maintenance, and revitalization efforts?

Questions such as these situate this study at the intersection of several linguistic subdisciplines (cf. §1.2). The following section seeks to integrate recommendations from each of these fields into the approach taken in analyzing linguistic variation here. Beginning with a discussion of the range of linguistic phenomena under consideration and possible methods for their representation (§4.2.1), later sections consider specific phonemic (§4.2.1.1), lexical (§4.2.1.2), phonological (§4.2.1.3), morphological (§4.2.1.4), and syntactic (§4.2.1.5) patterns in which variation between varieties of Mennonite Plautdietsch has been reported. Given the extent of such differences, the remainder of the chapter turns its attention to the problem of producing a linguistic resource (here, a documentary corpus) that offers adequate coverage of variation in these features (§4.2.2). The development of this resource in partnership with members of the Saskatchewan Valley Mennonite community through successive stages of design, recording, transcription, and annotation is detailed in the final section (§4.2.3), before concluding in a brief summary and look forward to the analysis undertaken in Chapter 5.

#### *4.2 Research design*

This study adopts a hybrid methodology to investigating linguistic variation, drawing on

perspectives and practices from each of the linguistic disciplines surveyed in Section 1.2 to plan the development of documentation that can be used to address the research questions raised above. Even at the earliest stages of planning, these fields have much to contribute to the direction and later conduct of research. The insights of both dialectology and sociolinguistics into sociodemographic factors that commonly condition the distribution of linguistic variation in and across communities (e.g., age, sex, class, place of birth and residence) encourage systematic attention to these features in the Saskatchewan Valley, as well. By the same token, methods from dialectometry, quantitative sociolinguistics, and corpus linguistics that make use of permanent documentation in analysis might contribute not only to discerning patterns in the information gathered, but also draw helpful attention to the process of developing such records in ways that are amenable to multiple scenarios of reuse. The approaches taken to the analysis of variation in each of these subfields might thus inform the present research, such that its outcomes are able to address the linguistic concerns reflected in the questions that opened this chapter.

As the literature in documentary linguistics has emphasized, there is often also potential in research involving smaller and endangered languages for such work to be relevant not only to linguistics as a discipline, but also to individuals and communities with an interest in their own language(s) (cf. Boas 2007, Himmelmann 2008: 345). Responsiveness to the interests of a range of stakeholders in the co-definition of research goals and practices, this literature argues, can contribute substantially to the overall usefulness of project outcomes not only for current linguistic research, but also for language activism, education, and revitalization. Given the goals of this study, particularly as reflected in the fourth research question raised above, it seems appropriate to seek balance between the general methodological recommendations of particular linguistic subfields and the specific priorities and interests of members of Saskatchewan Valley Mennonite Plautdietsch communities involved in the present research. With this in mind, guidance was sought from many individuals in the Saskatchewan Valley Mennonite community on how such ‘language work’ might best be pursued collaboratively here, and what forms its outcomes might take in order to reflect the interests of members of local communities.<sup>28</sup> As the

---

28 Particular acknowledgment is due to several individuals who provided critical perspectives on the potential for such research to be relevant to local language-related programs, and who contributed their considerable expertise with Mennonite language, history, and culture in the Saskatchewan Valley in guiding the direction of this study



author was raised in the Saskatchewan Valley in a largely Mennonite community, many of these individuals were acquainted with the author through kinship and family history, education, and/or participation in local Mennonite events and organizations. Several general recommendations emerged out of these discussions:

- The resulting materials should be usable in local language initiatives, such as beginners' courses in Mennonite Plautdietsch offered for adult language learners. For such programmes, few resources are available as supports for language learning, and few, if any, existing resources provide representation of Saskatchewan Valley Mennonite Plautdietsch varieties. Community involvement in this research, it was suggested, might contribute to more adequate representation of Plautdietsch as it is presently spoken in the area, and result in additional resources for language learners in the community;
- The final products of this research should be generally accessible so that both community members and others have the ability to consult and use the resulting resources in the long term. Although no previous linguistic research on Mennonite Plautdietsch has been conducted in the Saskatchewan Valley, the experience of some community members with past research in other academic disciplines was reportedly not always positive. Cases were shared in which cultural and historical information was collected in Saskatchewan Valley Mennonite communities by visiting researchers and later deposited in national and international archives, consequently limiting local access to such resources. It was emphasized that it was important to ensure that such situations did not happen again, and that all community contributions remain locally accessible;
- It was recommended that care be taken in the selection of appropriate research methods, as contributors' experience with spoken Plautdietsch may not always be reflected in their level of comfort with written forms of the language. Although virtually all individuals in the community are literate in English, an insistence on methods that involve written forms of Plautdietsch could limit the range of possible contributors. It was thus strongly encouraged that non-written means be considered to facilitate involvement in this project.

These discussions provided further motivation for considering research methods and instruments

---

as local advisors, among them Dick Braun, Kathy (Guenther) Braun, Jake Buhler, Leonard Doell, and Jack Driedger.

that reflected these concerns and that could result in resources that would be suitable for both the present investigation and for community language initiatives. Such conversations eventually led to the proposal of a Plautdietsch-language *Fibel* /ˈfi.bəl/ or primer. As described in Section 2.5, the traditional Russian Mennonite educational system supported literacy in Standard German by a progression of learning materials, beginning with a *Fibel* and proceeding through readings from the Bible and the Catechism. In contrast to the latter two sources, the introductory *Fibel* was often richly illustrated, and contained both individual words and sentences providing examples of particular orthographic conventions that often proceeded alphabetically through the sounds of the language. Adapting this kind of language learning resource to Plautdietsch thus presented an opportunity to make intentional connections to this important element of traditional Mennonite educational practices in the Saskatchewan Valley, and similarly to encourage the involvement of a wider range of contributors for whom this model was familiar and appealing.

From a linguistic perspective, a *Fibel* might also be viewed as a kind of documentary survey instrument, with a careful selection of words and sentences providing systematic representation of linguistic variables reported in other, related speech communities. In this respect, the *Fibel* resembles the use of standardized surveys in dialectology, or supplementary reading tasks and phonemic word lists in sociolinguistics. In both cases, these methods attempt to achieve consistent representation of linguistic features of interest across speakers, such that later comparisons of responses do not suffer from the sparseness that is often encountered with data from corpora of fully spontaneous speech (cf. Moisl 2009). As with such methods, a *Fibel* also has the advantage of being bounded in length, thus making it possible to involve more contributors from throughout the Saskatchewan Valley than would otherwise be feasible in the scope of this project. Accompanied by additional information about contributors and an appropriate selection of recording methods and annotation techniques, the process of developing a Plautdietsch-language *Fibel* could be seen as creating a specialized kind of documentation-based corpus—one which might later be mobilized for use in both linguistic analysis and community-based language initiatives (Nathan 2006).

While developing a Plautdietsch-language *Fibel* as both a community language resource and as a research instrument for the study of linguistic variation appeared to satisfy several of the aims of this research and the recommendations of community members, the selection of this

method was also not without attendant limitations, as discussed in Section 4.2.2. In order to assess the relative merits of this means of investigation, the following sections provide more detail on the linguistic design of the *Fibel*, beginning with the conception and representation of variation (§4.2.1) and proceeding to consider particular phonemic (§4.2.1.1), lexical (§4.2.1.2), phonological (§4.2.1.3), morphological (§4.2.1.4) and syntactic (§4.2.1.5) features reported to vary across Mennonite Plautdietsch communities. After surveying variation in these areas, the discussion returns to how these variables are to be incorporated into the *Fibel*, and how their representation might be improved through multiple rounds of piloting and revision with guidance from members of the Saskatchewan Valley Mennonite community.

#### 4.2.1 *Linguistic items and linguistic variables*

To incorporate linguistic variation coherently into the design of a Plautdietsch-language *Fibel*, it is first necessary to establish a framework within which such variation will be treated. This study adopts the sociolinguistic concept of a linguistic variable (cf. §1.2.1) as an abstraction over several linguistic features (*items*) considered to be variants in a single alternation (*variable*). In many cases, the assignment of items to a variable is relatively straightforward (e.g., with vowel quality differences between otherwise identical, synonymous words, as with Plautdietsch *kjlien* and *kjleen* ‘little’). In others, however, the decision of what should be treated as a single variable may be less clear. For instance, variation that occurs in specific phonological environments might be treated either by reference to individual lexical items or at a more abstract level (although with the concomitant risk of running afoul of lexical exceptions; cf. Cox, Driedger & Tucker 2013: 225 for one such example). Issues of appropriate granularity and abstraction such as this have notable consequences for how variables considered in the following sections are defined, and motivate careful attention to the documentation of particulars as an empirical safe-guard against unwarranted generalizations.

In computational terms, associating the features of an annotated corpus with particular items and variables is not difficult to accomplish (cf. Nagy & Meyerhoff 2013 for a recent example from quantitative sociolinguistics). A small program can be defined to accept a set of features extracted from a corpus (e.g., *moake* ‘to make’, as found in a corpus sentence) and a set of definitions that map aspects of these features onto items of particular variables (e.g.,

associating *moake* with a lack of pre-velar fronting in the diphthong /ɔa/ and with an *-e* infinitival ending). This program then gathers together all of the resulting items under their corresponding variables, presenting them as its final output. In this study, these definitions are given as Unicode-based regular expressions that are associated with particular items and variables. As an example, the following two definitions establish an association between a phonological variable (vREALIZATIONOAPREVELAR) and a number of features coded explicitly in the corpus (LXCook, LXCooked, LXDAYS, etc.). In the first definition, any of these features that contain /oə/ or /oː/ result in the variable vREALIZATIONOAPREVELAR being assigned the value ‘back’ (i.e., not fronted before velars). In the second definition, items that contain /eɔ/ or /øɔ/ cause the variable to have the value ‘front’ (i.e., fronted before velars):

```
vRealizationOaPreVelar (PHON): /.*(oə|oː).*/ --> "BACK" in lxCook,
lxCooked, lxDays, lXMake, lXOften, lXToday, W19IPA, W39IPA
vRealizationOaPreVelar (PHON): /.*(eɔ|øɔ).*/ --> "FRONT" in lxCook,
lxCooked, lxDays, lXMake, lXOften, lXToday, W19IPA, W39IPA
```

This system allows some flexibility in how variables are defined, at least insofar as the relevant patterns can be captured with regular expressions. This also permits the association of individual features with multiple variables, even when the coding of those variables may differ. For example, if some features are represented orthographically and others in the International Phonetic Alphabet, separate regular expressions can be defined to associate both sets of features with the same variable.

Importantly, this same procedure can be applied not only to corpus-derived items, but also to the contents of existing variables. This affords the potential to merge multiple, more concrete variables into larger, more abstract ‘macro-variables’, allowing potentially related phenomena to be considered at different levels of abstraction. Thus, with appropriate definitions, it is not only possible to treat the realization of homophonous *-e(n)* inflectional endings for third-person verbal subjects (e.g., *wi foahre(n)* ‘we’re driving’) and for plural nouns (e.g., *Bääare(n)* ‘berries’) as separate variables, but also to examine the viability of a single, more general pattern of variation in the realization of final *-e* or *-en* when these two variables are merged. In the definitions given below, for instance, separate morphological macro-variables are defined for verbal *-e(n)* (vENVERBAL), nominal *-e(n)* (vENNOMINAL), and other word-final *-e(n)* (vENOTHER)

endings by combining the values reported in several other, existing variables (e.g., for verbal *-e(n)*, by considering what endings are reported for third-person plural inflection (vEN3PL), infinitival endings (vENINF), and certain past participles (vENPASTPART)). These three macro-variables are then gathered together in a single, overarching macro-variable (vENGENERAL) that considers the general behaviour of *-e(n)* endings across all of these contexts.

```
vEnVerbal (MORPH):  /^(.*)$/ --> "\1" in vEn3PL, vEnInf, vEnPastPart
vEnNominal (MORPH): /^(.*)$/ --> "\1" in vEnNounSg, vEnNounPL
vEnOther (MORPH):  /^(.*)$/ --> "\1" in vEnOften, vEnWithout
vEnGeneral (MORPH): /^(.*)$/ --> "\1" in vEnNominal, vEnVerbal,
vEnOther
```

Definitions that establish macro-variables on the basis of other variables thus serve an important role in this analysis, providing the ability to vary the level of abstraction as necessary, with more abstract and generic variables being based directly on those with more concrete instantiations.

Even with such a mechanism in place to allow variables to be defined on the basis of corpus-derived data, there still remain questions as to the forms and extent of variation that should be considered for later analysis to be both circumspect and reasonably circumscribed. With the aim of this study being an initial investigation of linguistic differentiation in Saskatchewan Valley, and with no preceding linguistic investigations in these communities to guide the direction of research towards particular aspects of variation over others, an inclusive stance was taken towards what variation *has* been reported in the literature on Mennonite Plautdietsch and what can be inferred from available records of language use, rather than limiting the scope of variation prematurely. This position is much in line with the suggestions of the Tyneside Linguistic Survey and Dorian (2010) mentioned in Section 1.2.1, both of which attempt to avoid an overly narrow selection of variables that might predetermine the results of ensuing analyses. Likewise, bearing in mind the criticisms of traditional dialectology in its narrow concentration on lexical and morphophonological variation, it is also reasonable to aim for representation of variability in multiple aspects of linguistic organization (insofar as this is possible with the chosen methods) to achieve a more balanced representation of variation across the language. While this potentially implies a larger number of variables to be considered, one can seek to draw on contemporary quantitative and statistical methods in use elsewhere in linguistics to assist in rendering larger-scale, multivariate analysis tractable (cf. §5.2).

This leaves the question of how relevant variables are to be identified and brought into the design of the *Fibel*. In some cases, it is possible to draw on reports of variation in other Mennonite Plautdietsch speech communities (such as those encountered in Chapter 3) to arrive at set of phenomena in which some degree of linguistic differentiation might be found in Saskatchewan Valley communities, as well. In other cases, given the underdocumented state of linguistic practices in the Saskatchewan Valley, it is not surprising that some variation should emerge in the course of developing the *Fibel* without this having been planned in its design. Both the former, better documented variation and the latter, underdocumented variation are summarized in Appendix C, which provides a more detailed overview of the linguistic items discussed below. The following sections concentrate on how such variation is reflected in different aspects of linguistic organization and how this might be represented as variables, paving the way for these elements to be integrated into the design of the *Fibel* discussed in Section 4.2.2.

#### *4.2.1.1 Phonemic inventory*

When considering what elements of linguistic organization should be consistently present in an initial survey of variation across varieties of Plautdietsch, the basic set of phonemic contrasts presents itself as one relevant target of investigation. Although overall phonemic inventories have generally received less attention in traditional dialectological research than specific instances of phonological and morphological variation, the representation of major phonemic contrasts in controlled phonological environments is increasingly seen relevant not only to the baseline phonetic description of linguistic varieties, but also as an important resource for sociophonetic investigations of subsymbolic variation (Ladefoged 2003, di Paolo & Yaeger-Dror 2010). With Plautdietsch varieties differing considerably in their reported phoneme inventories, this section draws comparisons between published analyses, with the aim of arriving at an overall set of phonemic categories and their major allophonic variants that provides systematic representation of all significant contrasts reported in the literature.

There is general agreement across phonemic analyses of Plautdietsch on a common inventory of consonants, albeit with notable differences in both the treatment of certain sounds as allophones or independent phonemes and in the phonetic targets reported for some phonemic

categories. Table 2 presents one such analysis of a variety of Mennonite Plautdietsch spoken in the Saskatchewan Valley that serves to illustrate the major phonemic contrasts of the language (after Cox, Driedger & Tucker 2013: 222):

	Bilabial	Labio-dental	Alveolar	Palato-alveolar	Palatal	Velar	Glottal				
Plosive	p	b	t	d	k <sup>j</sup>	g <sup>j</sup>	(?)				
Nasal		m		n		ɲ					
Affricate			ʈʂ		tʃ						
Fricative		f	v	s	z	ʃ	ʒ	ç	x	(ɣ)	h
Trill				r							
Approximant						j					
Lateral approximant				l		l <sup>j</sup>					

Table 2. Plautdietsch consonant phonemes.

While some phonemic categories are only attested in a limited range of lexical items, and thus not recognized in all analyses (e.g., /l<sup>j</sup>/, which appears only in a handful of words in a limited number of phonotactic environments; cf. Kanakin & Wall 1994: 14), the general structure of the consonantal phoneme inventory is uncontroversial. Rather, what differences do exist between analyses appear mainly in the status of particular segments as phonemes or allophones, and in the attestation of particular allophones across varieties of Plautdietsch:

- Several authors suggest that [ɣ] could be treated as an allophone of /g/, appearing in intervocalic and final positions (Loewen 1998: 135) and, in some varieties, word-initially before back vowels (Jedig 1966, Klassen 1969: 29, Nyman 1978: 53);
- Some authors (e.g., Kanakin & Wall 1994: 15) argue that [h] and [x] represent pre- and post-vocalic allophones of a single phoneme. This observation may hold for native Plautdietsch lexemes, but not for Slavic borrowings found in some varieties (e.g. *xotj* [xɔc] ‘even, at least, even though’, *Chomut* [xɔmɔt] ‘(horse) collar’; see Siemens 2012: 209 and Epp 1996: 81–82 for examples), motivating the treatment of these segments as separate phonemes here;
- Some varieties of Plautdietsch reportedly do not distinguish /ʈʂ/ and /s/ word-initially and in stressed syllable onsets (Siemens 2012: 112–113). This has led some authors to

treat /tʃ/ as two segments, rather than as a single affricate. In a similar way, some authors (e.g., Kanakin & Wall 1994, Siemens 2012) do not treat /tʃ/ as a phoneme in its own right, but rather as a combination of /t/ and /ʃ/.

- It is not uncommon for analyses to omit the glottal stop [ʔ] from the overall phoneme inventory, arguing that this segment appears predictably as an epenthetic onset in syllables that would otherwise lack one (cf. Cox, Driedger & Tucker 2013, but see Loewen 1998: 131 for arguments against this position).
- The typologically unusual palatalized stops represented as /kʲ/ and /gʲ/ in Table 2 are a topic of much commentary in the literature on Mennonite Plautdietsch, particularly concerning their origins and phonetic realization in contemporary varieties (cf. Baerg 1960, Moelleken 1966, Nyman 1978: 52, Reimer, Reimer, & Thiessen 1983, Kanakin & Wall 1994, Siemens 2012: 93–98, a.o.). For the present purposes, it is enough to observe two separate categories of sounds here, the realization of which may vary between varieties.

These observations notwithstanding, the consonantal phoneme inventory of Plautdietsch appears relatively stable across varieties, although certainly admitting variation in the realization of these categories. By comparison, considerably more variation is encountered in phonemic analyses of Plautdietsch monophthongs. Several analyses are summarized in Table 3, which compares the reported realizations of a maximal set of phonemic contrasts attested in varieties of Plautdietsch.



<i>Study</i>	<i>Variety</i>	/i/	/ɪ/	/e/	/ɛ/	/a/	/ə/	/o/	/ɔ/	/u/	/ʊ/
Baerg (1960)	USA (M1870)	î [i]	i [ɪ]	/ê/, /ē/ [e], [ɛ:]	e [ɛ]	a [ʌ, a, a:]	ə [i, ə]	/ô/, /ō/ [o], [õ]	o [ɔ̃, <ɔ̃]	û [u, u]	u [u <sup>1</sup> , ʊ]
Cox et al. (2013)	Canada (C1870)	/i:/	/ɪ/	/e/	/ɛ/	/a/	/ə/	/o/	/ɔ/	/y/	/ʊ/
Epp (1996)	Canada (M1870)	i, ie [i]	i [ɪ]	ä, e, eh [e]	ä, e [ɛ]	a [a]	e [ə]	o, oh [o]	o [ʌ]	u, uh [u]	u [ʊ]
Goerzen (1950)	Canada (<M1920)	ie, i	i	e:	ê	a, aa	ê	ô	o	u:	û
Kanakin & Wall (1994)	Russia (<C)	/i:/	/ɪ/	/e:/, iə, i:/	/ɛ/	/a:/	(/ə/)	/o:/, u:/	/ɔ/	/y:/	/ʊ/
Klassen (1969)	Russia (<C/M)	/i:/	/ɪ/	/e:/	/ɛ/	/a/, /a:/	/ə/	/o:/	/ɔ/	/y:/	/ʊ/
Lehn (1957)	USA (C1920)	ii [i:]	i [ɪ]	ei [e:]	e [ɛ]	a [a:, a]	ə [i, ə]	ou [o:]	o [ɔ]	uu [u:]	u [ʊ]
Loewen (1996)	Canada (C1920)	ie	i	ä	e	a	ê	oo	o	ũ	u
McCaffery (2008)	USA (M1870)	ie	i	ää	e	a	e	o	o	uu	u
Mierau (1964)	Ukraine (C1920)	ii	i	ee	e	a	i	oo	o	uu	u
Moelleken (1966)	Mexico (C1870)	/i/	/ɪ/	/e/	/ɛ/	/a/	/ə/	/o/	/ɔ/	/y/	/ʊ/
Moelleken (1967)	Canada (1920/40)	/i/	/ɪ/	/e/	/ə/	/a/	/ə/	/o/	/ɔ/	/u:, ʌ:, y:/	/ʊ/
Moelleken (1972)	Canada (1920)	/i:/	/ɪ/	/e:/	/ɛ/	/a/	/ə/	/o:/	/ɔ/	/u:, ʌ:, y:/	/ʊ/
Naiditsch (2001)	(n/a)	/i:/	/ɪ/	/e:/	/ɛ/	/a:/		/o:/	/ɔ, ɔ:/	/u:, y:/	/ʊ/
Neufeld (2000a)	USA (M1870)	ie	i	ä	e	a	e	o	o	uu	u
Nieuweboer (1998)	Russia (<C/M)	/i:/	/ɪ/	/e:/	/ɛ/	/a/	/ɛ/	/o:/	/ɔ/	/y:/	/ʌ/
Nyman (1978)	Russia (<C/M)	/i:/	/ɪ/	/e:/	/ɛ/	/a/	/ə/	/o:/, ɔʊ/	/ɔ/	/u:, y:/	/ʊ/
Quiring (1928)	Ukraine (C1920)	î	i	ē	e	a, â	ə	ō	o	û	u
Reimer et al. (1983)	Canada (C/M)	ie [i]	i	ä, e [e]	e [ɛ]	a [ʌ, a]	e [ə]	o [o]	o [ɔ]	u [u]	u [ʊ]

<i>Study</i>	<i>Variety</i>	/i/	/ɪ/	/e/	/ɛ/	/a/	/ə/	/o/	/ɔ/	/u/	/ʊ/
Rempel (1995)	Canada (C1870)	ie	i	ä	e	a	(e)	o	o	ü	u
Thiessen (2003)	Canada (C1920)	ie	i	ä, e	e	a	e	o	o	ü	u
Zacharias (2009)	Canada (C1870)	ie	i	ä	e	a	(e)	o	o	u	u

*Table 3.* Phonemic analyses of Plautdietsch monophthongs. Varieties are listed by country, period of emigration from Russia or Ukraine, and colony of origin (C = Chortitza Colony, M = Molochnaya Colony, <C/M = Russian-Ukrainian daughter colony of Chortitza/Molochnaya).

As with the consonantal phoneme inventory, individual studies of these vowels differ substantially in their treatment of particular segments:

- Although this summary treats the contrasts between monophthongs as relating primarily to vowel quality, several authors point to length distinctions between vowels that may offer another valid means of classification (Mierau 1964, Klassen 1969, Moelleken 1972, Nieuweboer 1998, Naiditsch 2001; see Cox, Driedger & Tucker 2013 for discussion);
- Most Plautdietsch orthographies use the same grapheme for /o/ and /ɔ/ (or /o:/ and /o/, depending on the analysis), employing other conventions to capture the phonemic distinction between these vowels (cf. Rempel 1995, Neufeld 2000a);
- Several authors omit schwa from the phonemic inventory altogether, sometimes treating it as an unstressed allophone of /ɛ/, instead (cf. Kanakin & Wall 1994);
- Perhaps the most prominent variation in monophthong inventories between Plautdietsch varieties lies in the realization of /u/. Moelleken (1967: 245–246) treats this feature in his study of *Russländer* Mennonite Plautdietsch varieties in British Columbia, relating its distribution to the Ukrainian colony with which speakers were associated. Moelleken reports that Chortitza speakers and those from mixed daughter colonies maintained only the realization [y:], while Molochnaya speakers varied between [u:], [ʊ:], and [y:], a conclusion consonant with the reports of Baerg (1960) and Nyman (1993).

Considering the monophthongal inventory as a whole, the greatest number of distinct categories is reported by Baerg (1960), who distinguishes two separate monophthong phonemes in the /e/ and /o/ spaces. Unlike the other studies summarized here, Baerg reports the phonemes she transcribes as *ê* /e/ and *ô* /o/ to be monophthongs, rather than the diphthongs \**əɪ* and \**əʊ* found in most other studies. These two phonemes aside, the remaining categories appear to be common across all surveyed varieties, and thus potentially suitable as a basis for comparisons of phonemic realization across varieties.

In contrast to the consonantal and monophthongal phonemes, diphthongs present a much more diverse picture, with multiple diachronic paths of development leading to different numbers of contrasts in contemporary varieties. The resulting lack of consensus on the number of contrasting diphthongs and their phonetic realizations is evident in the summary in Table 4.<sup>29</sup>

---

<sup>29</sup> Also evident is the marginal status of several of these contrasts, especially /uɪ/, which appears in only a handful

<i>Study</i>	<i>Variety</i>	/ia/	/ea/	/ɛa/	/ɛɪ/	/əɪ/	/əʊ/	/aʊ/	/ɔa/	/oa/	/ua/	/uɪ/
Baerg (1960)	USA (M1870)	îe [i̇]	êe [ė]	ēə [Ė]	ei [ɛ̇i]	(ê)	(ô)	ou [ɔ̇u]	ōə [ȯ]	ôə [ɔ̇]	/ûə/ [u̇]	
Cox et al. (2013)	Canada (C1870)	/ia/	(/əɪa/)	/ea/	/ɛɪ/	/əɪ/	/əʊ/	/aʊ/	/oa/; /eɔ/	/ua/; /ya/	/ya/	/uɪ/
Epp (1996)	Canada (M1870)	ia	ea	āa	ei	ee	oo	au	oa	ua, ooa	ua	uj
Goerzen (1950)	Canada (<M1920)	ia [i:ë]	ea [e:ië]	āa [e:ë]	ei [ée]	ee	oo	au [áo]	ôa	oa [ou:ë]	ua [ú:ë]	
Kanakin & Wall (1994)	Russia (<C)	/iə, i:/	/iə, i:, ɛɪ/	/ɛɪ, ɛ:/	/ɛɪ, ɛ:/	/əɪ, ɔɪ/	/əʊ/	/ɔu, ɔ:/	/ɔɪ, ɔ:/	/uə, u:, o:/	/yə/	
Klassen (1969)	Russia (<C/M)	ia	(eia)	ea [ɛa]	äe [ɛë]	eī	ou	au [ɔō]	oa	oua	üa	
Lehn (1957)	USA (C1920)	(ii+a)	(ei+a, əi+a)	(e+a)	ai [aɛ]	əi [əi]	əu [əu]	au [aʊ]	(ou+a)	(əu+a)	(uu+a)	
Loewen (1996)	Canada (C1920)	iea	(eea)	āa	ei	ee	ou	au	oa; ða	ua, oua; üa	üa	
McCaffery (2008)	USA (M1870)	ia, iee	ea	āa	ei	ee	oo	au	oa	ua	?	
Mierau (1964)	Ukraine (C1920)	(ii)	(ai)	(ee)	ei	ai	au	ou	(oo)	(uu)	?	
Moelleken (1966)	Mexico (C1870)	?	/ia/	/ɛa/	/æ/	/ɔɪ/	/əʊ/	/ɔ/	/ɔa/; /ɛa/	/ua/; /ya/	?	
Moelleken (1967)	Canada (1920/40)	?	/ea, əɪ, ɪa/	/ɛa, e:/	/ɛɪ/	/əɪ, e:, ɔɪ/	/əʊ/	/ɔʊ/	/ɔa, o:/; /œa/	/ua, oa, əʊ/; /øa/	?	
Moelleken (1972)	Canada (1920)	?	/ea/	/ɛa/	/æɪ/	/əɪ/	/əʊ/	/ɔʊ/	/ɔa/	/oa/	?	/oɪ/
Naiditsch (2001)	(n/a)	?	/iə/	/eɐ/	/ɛi/	/ɔi/	/əʊ, ø:/	/ɔʊ, ɔ:/	/oɐ/	/uɐ/	/yɐ/	
Neufeld (2000a)	USA (M1870)	ia	ea	āa	ei	ee	oo	au	oa	ua	ua	
Nieuweboer (1998)	Russia (<C/M)	/iə/	/iə/	/eɐ/	/ɛɪ/	/əɪ/	/əʊ/	/ɔ:/	/oɐ/	/uɐ/	/yɐ/	

of lexical items (e.g., *fuj* /fuɪ/ ‘yuck!’) and thus often escapes attention in treatments of diphthongs in Mennonite Plautdietsch; cf. Cox, Driedger & Tucker (2013: 225). In general, it should be noted that not all scholars analyze these vowel sequences as phonemes in their own right; see Lehn (1957) and Mierau (1964) for analyses that treat several of these purported diphthongs as allophones of long monophthongs, and Naiditsch (2001: 248) for arguments against the tenability of this position.

<i>Study</i>	<i>Variety</i>	/ia/	/ea/	/ɛa/	/ɛɪ/	/əɪ/	/əʊ/	/aʊ/	/ɔa/	/oa/	/ua/	/uɪ/
Nyman (1978)	Russia (<C/M)	/ea/	(/əɪa/)	/ɛa/	/æɪ/	/əɪ/	/əʊ/	/ɔʊ/	/ɔa/	/oa, o:, u:/	(n/a)	oɪ
Quiring (1928)	Ukraine (C1920)	îə	êə	ēə	ei	(ê)	(ô)	au	ōə, ȯo; ōo	ôə; ȯo	ûə	
Reimer et al. (1983)	Canada (C/M)	ia	ea	äa	ei	ee	oo	au	oa	ua	ua	
Rempel (1995)	Canada (C1870)	ia	ea	äa	ei	ee	oo	au	oa	ua; üa	üa	uj
Thiessen (2003)	Canada (C1920)	ia	ea	äa	ei	ee	oo	au	oa	oa, ua	üa	
Zacharias (2009)	Canada (C1870)	ia, iee	ia, iee	äa	ei	ee	oo	au	oa	ua	ua	

*Table 4.* Phonemic analyses of Plautdietsch diphthongs. Varieties are listed by country, period of emigration from Russia or Ukraine, and colony of origin (C = Chortitza Colony, M = Molochnaya Colony, <C/M = Russian-Ukrainian daughter colony of Chortitza/Molochnaya). Where noted, distinctive pre-velar allophones of /ɔa/ and /oa/ are given after a semicolon in the corresponding column.

While there are thus considerable differences in the inventories of contrasting diphthongs in present-day Mennonite Plautdietsch varieties, this variation is not without coherent patterning. In the front diphthongs, some varieties (e.g., Goerzen 1950, Baerg 1960, Epp 1996, Thiessen 2003) maintain a three-way contrast between *\*ia*, *\*ea*, and *\*äa*, while others (e.g., Nyman 1978, Nieuweboer 1998, Zacharias 2009, Cox, Driedger & Tucker 2013) do not generally distinguish *\*ia* and *\*ea*, as seen in Table 5 below. Even in the latter varieties, however, some lexical items (e.g., *vea* ‘four’, *Bea* ‘beer’, *mea* ‘more’) still reflect this distinction, as illustrated in Table 6 (see Cox, Driedger & Tucker 2013: 226 for further discussion).

	<i>fiare(n)</i> ‘celebrate’	<i>feahre(n)</i> ‘carry’	<i>vääre(n)</i> ‘in front’
Epp (1996)	/fiəɾə/	/fəiəɾə/	/fəəɾə/
Zacharias (2009)		/fiarən/	/fəərən/

Table 5. Front diphthong contrasts in Plautdietsch varieties.

	<i>Fia</i> ‘fire’	<i>vea</i> ‘four’	<i>Fäah</i> ‘Fehr’ (name)
Epp (1996)	/fia/	/fəia/	/fea/
Zacharias (2009)	/fia/	/fəia/	/fea/

Table 6. Lexical exceptions to front diphthong contrasts in Plautdietsch varieties.

Similar variation is noted in back diphthongs, albeit with an additional complication. In some varieties of Plautdietsch, *\*ɔa* and *\*oa* have front rounded allophones *\*œa* and *\*øa* before velar consonants (cf. Moelleken 1967). In the most conservative of these varieties (e.g., Quiring 1928), there are five distinctive allophones of the three back diphthongs, as seen in Table 7. In other varieties, however, several of these contrasts have been lost, leading to different configurations of these sounds (cf. Rempel 1995; Loewen 1996, 1998; Zacharias 2009). In other varieties where no distinctive pre-velar allophones exist, all three contrasting back diphthongs may remain distinct (e.g. Goerzen 1950, Nieuweboer 1998), or additional mergers may have taken place, leaving a smaller number of contrasts (e.g., Epp 1996, but see Epp 1996: 35–37 for several lexical exceptions).<sup>30</sup>

30 The set of back diphthong allophones and phonemic contrasts presented here should not be taken to be exhaustive; there is some evidence in the literature of further phonological variation. For instance, Moelleken (1967) reports speakers from Gnadenfeld (Molochnaya) as having allophonic [o:] in free variation with [ɔa] for pre-velar *\*ɔa* (*\*œa*), and allophonic [əʊ] in free variation with [oa] for pre-velar *\*oa* (*\*øa*)—realizations which

	Quiring (1928)	Loewen (1996)	Zacharias (2009)	Rempel (1995)	Goerzen (1950)	Epp (1996)
*ɔa ( <i>woah</i> ‘true’)	ɔa	ɔa [oa]	ɔa [oa]	ɔa [oa]	ɔa [oa]	ɔa [oa]
*œa ( <i>Woag</i> ‘dare’)	œa	œa [øa]	œa [øa]	œa [øa]	œa [øa]	œa [øa]
*oa ( <i>wua</i> ‘where?’)	oa	oa [ua~əua]	oa [ua]	oa [ua]	oa [əua]	oa [ua]
*øa ( <i>wuag</i> ‘weighed’)	øa	ua [ya]	ua [ya]	ua [ya]	ua [ua]	ua [ua]
*ua ( <i>Bua</i> ‘builder’)	ua [ya]					

Table 7. Back diphthong contrasts in Plautdietsch varieties.

Differences in the diphthong inventories of Plautdietsch varieties are not limited to systemic contrasts such as these. Even where the phoneme categories of two varieties are essentially the same in terms of their overall organization, differences may still exist in the realization of individual categories. This is well exemplified by western Siberian Plautdietsch, where an extensive reconfiguration of vowel targets is attested, even while the phonemic categories themselves remain stable (e.g., Kanakin & Wall 1994, Nieuweboer 1998, Naiditsch 2001, but not Klassen 1969). The range of realizations of each phoneme presents another area in which differences between speakers and varieties might be found, as would be expected from the burgeoning literature in sociophonetics. Adequate documentation of such varieties should thus be planned in such a way as to enable further research along such lines, even if questions of mean phonetic realization fall outside of the immediate scope of investigation here.

Finally, it is also possible to treat several triphthongs as distinct phonemes, although their analysis is less clear-cut than those of either monophthongs or diphthongs due to their relative infrequency and apparent reanalysis as bisyllabic sequences in some varieties. Of the triphthongs that have been reported in contemporary varieties of Plautdietsch, only two, [əia] and [əua], end in a non-glide. The former, [əia], is attested in some varieties as a variant of the diphthong \*ea (e.g., *Bea* /bəia/ ‘beer’, *mea* /məia/ ‘more’) and could be treated as a triphthong, even if its phonological patterning is essentially the same as other diphthongs (cf. Cox, Driedger & Tucker 2013: 226). Similarly, the triphthong [əua] is attested as a variant of the diphthong \*oa (e.g., *Dooa* /dəua/ ‘gate, gateway’, *Ooah* /əua/ ‘ear’; Loewen 1998: 136–137, Epp 1996: 35–37), although this contrast appears to be marginal in even the few varieties where it is attested. The remaining attested triphthongs all end in an off-glide /ɪ/ or /j/ and could be treated as

---

are not reported for Mennonite Plautdietsch speakers from other villages or settlements.

diphthongs followed by a coda consonant, rather than as separate triphthong phonemes as such. Whatever their phonemic status, several of these sequences are relatively common in the Plautdietsch lexicon and arguably deserve attention when considering the overall patterning of sounds in the language. These include \**äaj* (e.g., *däaj* ‘considerably’), \**ej* (e.g., *Breaj* ‘scalding hot water’), \**auj* (e.g., *Krauj* ‘crow’), and \**ɔaj* (e.g., *Boaj* ‘mountains, hills’). A final triphthong, \**ooj*, is scantily attested, appearing only in a handful of lexical items that are not found in all varieties (e.g., *looj* ‘reluctant’, *mooj* ‘comfortable, warm; lazy’; cf. Thiessen 2006).

Drawing on these observations, it is possible to construct a consensus analysis of the maximal phonemic inventory found across contemporary Mennonite Plautdietsch varieties, making allowances for commonly reported instances of allophonic variation. This includes 27 consonant phonemes (plus an additional instance of /g/ to provide representation of a possible voiced velar fricative allophone in intervocalic position), ten monophthongs, eleven diphthongs (plus two instances of the mid-back diphthongs in pre-velar environments, in order to determine the presence or absence of fronted allophones), and at least four possible triphthongs. This results in a total of 55 phonemes and allophones that require separate consideration in comparisons of the overall phonemic inventories of Mennonite Plautdietsch varieties. Table 8 presents each of these segments with their orthographic representation and phonemic value, as well as a lexical item (numbered W01 to W55) in which the segment is instantiated. For the purposes of consistent cross-varietal phonological comparison, these lexical items were selected to be common across varieties of Mennonite Plautdietsch and not prone to lexical variation. This approach sought to avoid situations in which the associated word would be unfamiliar to speakers or could easily be substituted for another, semantically related form that was nevertheless phonetically distant.<sup>31</sup> In addition, to allow for future systematic analyses of the phonetic features of both the consonantal and vocalic inventories, care was taken to select lexical items which provided as consistent a phonetic environment as possible for the target segments. Wherever possible, vowel contrasts were illustrated in word-initial stressed syllables following a voiced bilabial (either /b/ or /v/) and preceding an alveolar stop (either /t/ or /d/). Similarly, for the consonantal contrasts, lexical items were preferred in which the relevant segment appeared

---

31 See also Section 4.2.2.2 on additional methods employed in the development of elicitation materials to restrict the scope of reference to the intended target words in these cases.



word-initially in a stressed syllable before /o/ except where lexical or phonotactic restrictions prevented this (following Cox, Driedger & Tucker 2013; cf. Ladefoged 2003).

W01	a	/a/	<i>Bad</i> /bad/ 'bed'	W28	kj	/ki/	<i>Kjoasche(n)</i> /kjoʌʃə(n)/ 'cherries'
W02	ä	/e/	<i>bädt</i> /bet/ 'prays'	W29	l	/l/	<i>lot</i> /lot/ 'late'
W03	äa	/εa/	<i>Bäa</i> /bea/ 'berry'	W30	lj	/li/	<i>Eelj</i> /ɛli/ 'oil'
W04	äaj	/εaɪ/	<i>Wäaj</i> /vεaɪ/ 'roads'	W31	m	/m/	<i>Mon(d)</i> /mon(t)/ 'moon'
W05	au	/aʊ/	<i>Bauss</i> /baʊs/ 'boss'	W32	n	/n/	<i>Noba</i> /noba/ 'neighbour'
W06	auj	/aʊɪ/	<i>Krauj</i> /kraʊj/ 'crow'	W33	ng	/ŋ/	<i>Hunga</i> /hʊŋa/ 'hunger'
W07	b	/b/	<i>Boa</i> /boʌ/ 'bear'	W34	nj	/ɲ/	<i>Kjinja</i> /kiɲa/ 'children'
W08	ch	/x/	<i>Dag</i> /dax/ 'day'	W35	o	/ɔ/	<i>Botta</i> /bɔta/ 'butter'
W09	d	/d/	<i>Dola</i> /dola/ 'dollar'	W36	o	/o/	<i>Wota</i> /vota/ 'water'
W10	e	/ə/	<i>besied</i> /bəzid/ 'beside'	W37	oa	/ɔa/	<i>Boat</i> /boʌt/ 'beard'
W11	e	/ε/	<i>betta</i> /beta/ 'bitter'	W38	oaj	/ɔaɪ/	<i>Boaj</i> /boʌɪ/ 'mountains'
W12	ea	/ea/	<i>Beakja</i> /beakja/ 'books'	W39	öa	/œa/	<i>Woage(n)</i> /vɔage(n)/ 'wagon'
W13	ea	/əɪa/	<i>vea</i> /fəɪa/ 'four'	W40	oo	/əʊ/	<i>Foot</i> /fəʊt/ 'foot'
W14	eaj	/eaɪ/	<i>Kjeaj</i> /kieaɪ/ 'cows'	W41	p	/p/	<i>Poa</i> /pɔa/ 'pair'
W15	ee	/əɪ/	<i>Beete(n)</i> /bɛɪtə(n)/ 'beets'	W42	r	/r/	<i>Rot</i> /rot/ 'advice'
W16	ei	/εɪ/	<i>Weit(e)</i> /vɛɪt(ə)/ 'wheat'	W43	s	/z/	<i>Sot</i> /zot/ 'seed'
W17	f	/f/	<i>Foahra</i> /fɔara/ 'driver'	W44	ss	/s/	<i>Massa</i> /masa/ 'knife'
W18	g	/g/	<i>Goade(n)</i> /gɔadə(n)/ 'garden'	W45	sch	/ʃ/	<i>schoap</i> /ʃɔap/ 'sharp'
W19	g	/ɣ/	<i>Foagel</i> /fɔagəl/ 'bird'	W46	t	/t/	<i>tohm</i> /tom/ 'tame'
W20	gj	/gɨ/	<i>Migje(n)</i> /mɨgɨə(n)/ 'mosquitos'	W47	ts	/ʦ/	<i>Zocka</i> /ʦɔka/ 'sugar'
W21	h	/h/	<i>Hoat</i> /hɔat/ 'heart'	W48	tsch	/ʧ/	<i>Dietsch</i> /diʧ/ 'German'
W22	i	/ɪ/	<i>witt</i> /vɪt/ 'white'	W49	u	/ʊ/	<i>Buck</i> /bʊk/ 'stomach'

W23	ia	/ia/	<i>hia</i> /hia/ 'here'	W50	ü	/u/	<i>Huus</i> /hus/ 'house'
W24	ie	/i/	<i>wiet</i> /vit/ 'far'	W51	ua	/oa/	<i>Wuat</i> /voat/ 'word'
W25	j	/j/	<i>Joah</i> /jɔa/ 'year'	W52	üa	/øa/	<i>Buak</i> /boak/ 'book'
W26	jch	/ç/	<i>Laicha</i> /laça/ 'holes'	W53	uj	/ʊi/	<i>fuj</i> /foi/ 'phooey!'
W27	k	/k/	<i>Koa</i> /kɔa/ 'car'	W54	w	/v/	<i>woat</i> /vɔat/ 'will (be)'
				W55	zh	/ʒ/	<i>Bockelzhann</i> /bɔkəl'ʒan/ 'tomato'

Table 8. Phonemic categories and allophones of Plautdietsch.

Although this list attempts to provide systematic coverage of all major phonemic distinctions and common allophonic variants attested in contemporary Mennonite Plautdietsch varieties in a form amenable to cross-dialectal comparison, there are nevertheless limitations to this approach. In particular, with this list being based on contemporary Mennonite Plautdietsch varieties, it is possible that comprehensive comparisons with other, more distantly related varieties (e.g., non-Mennonite forms of Plautdietsch) may be more difficult. While such varieties are admittedly not the central focus of this study, one possible alternative is outlined in Siemens (2012), who draws on the work of Niebaum (1985, reproduced in Niebaum 2000) to consider the maximal set of vowel distinctions attested historically in all Low German varieties. Siemens is thus able to select contemporary lexical items that represent each of these historical contrasts, even if they are no longer distinguished consistently in Mennonite Plautdietsch. A combination of this more diachronically oriented approach, together with attention to possibly idiosyncratic developments found in present-day Plautdietsch varieties, may afford another way of proceeding with systematic comparisons that extend beyond Mennonite Plautdietsch proper.

#### 4.2.1.2 Lexical variables

Differences related to features of the lexicon present another prominent way in which varieties may differ from one another (cf. §1.2.2). It is possible to distinguish several forms of lexical variation. Perhaps most obviously, there may be differences between varieties in the conventional choice of lexical items for particular concepts, with or without any phonological

resemblances between such forms. This is the case with Plautdietsch terms for ‘store’, for example, with some speakers favouring the term *Laufdje* /laʊftʃə/ (< Rus./Ukr. лавка /lavka/ ‘(general) store’; cf. Epp in Rempel 1995: v), others *Lode(n)* (cognate with Standard German *Laden*), and still others *Stua* /ʃtua/ (< Eng. *store*). In other, less common cases, lexical variation may relate to the inflectional class to which a lexeme is assigned. All varieties of Plautdietsch share the term *Voagel/Vöagel* ‘bird’, for instance, but differ in treating it either as a neuter (*daut Voagel/Vöagel*) or masculine (*de Voagel/Vöagel*) noun (Dyck 1964: 67, Nyman 1997: 267). These kinds of variation in either the selection of lexical items or their assignment to inflectional classes are grouped together here under the cover term ‘lexical variation’, with Table 9 providing an overview of the lexical variables considered in this study. (Source item labels in this table refer either to the specific features listed in Appendix C or to the phonemic variables given in Table 8 above).

<i>Variable</i>	<i>Variants</i>	<i>Source Items</i>
‘a (nom. m./n., reduced)’ (vMASCNOMAREduced)	<i>e</i> , ‘ <i>en</i>	LXMASCNEUTA
‘am’ (vAM)	<i>se</i> , <i>senn</i> , <i>si</i>	LXAM
‘are’ (vARE)	<i>sen</i> , <i>send</i> , <i>senne</i>	LXARE
‘aunts’ (vAUNTS)	<i>Mumms</i> , <i>Tauntes</i>	LXAUNTS
‘became (sg.)’ (vBECAME)	<i>word</i> , <i>wort</i>	LXBECAMESG
‘because’ (vBECAUSE)	<i>because</i> , <i>deswäajen(s)</i> ( <i>daut</i> ), <i>doawäajen(s)</i> ( <i>daut</i> ), <i>wäajen(s)</i> ( <i>daut</i> ), <i>weens/winjs</i> , <i>weil</i> , <i>wiel</i> , <i>wiels</i> , <i>wielt</i>	LXBECAUSE
‘been’ (vBEEN)	<i>jewas(t)</i> , <i>jewäse(n)</i>	LXBEEN
‘beside’ (vBESIDE)	<i>besied</i> , <i>bersied</i> , <i>bisied</i>	W10
‘between’ (vBETWEEN)	<i>teschen</i> , <i>tweschen</i>	LXBETWEEN
‘bird (gender)’ (vBIRDGENDER)	MASC., NEUT.	LXBIRDGENDER
‘corn, maize’ (vCORN)	<i>Korn</i> , <i>Kuckurus</i>	LXCORN
‘days’ (vDAYS)	<i>Do(a)g</i> , <i>Do(a)ge</i>	LXDAYS
‘down, off’ (vDOWN)	<i>(e)rauf</i> , <i>(e)raufa</i>	LXDOWN
‘early’ (vEARLY)	<i>fräh</i> , <i>tiedig</i>	LXEARLY
‘eaten’ (vEATEN)	<i>jeäte(n)</i> , <i>jejäte(n)</i>	LXEATEN
‘English’ (vENGLISH)	<i>Engelsch</i> , <i>Englisch</i>	LXENGLISH
‘farmer’ (vFARMER)	<i>Foa(r)ma</i> , <i>Bua</i>	LXFARMER
‘George’ (vGEORGE)	<i>Jeat</i> , <i>Jorg</i>	LXGEORGE
‘girls’ (vGIRLS)	<i>Me(r)jalle(n)</i> , <i>Me(r)jalles</i> , <i>Mäakjes</i>	LXGIRLS
‘grandmother’ (vGRANDMOTHER)	<i>Groos(ma(u)</i> , <i>-mam(e)</i> , <i>-mutta</i> ), <i>Grootmutta(r)(tje)</i>	LXGRANDMOTHER

<i>Variable</i>	<i>Variants</i>	<i>Source Items</i>
'have (unreduced)' (vHAVEFULL)	<i>habe(n), hawe(n)</i>	LXHAVEINF, LXHAVEAUXPL, LXHAVELEXPL
'have (reduced)' (vHAVEREDUCED)	<i>ha', ha'n, ha'nen</i>	LXHAVEINF, LXHAVEAUXPL, LXHAVELEXPL
'her (dat.)' (vDATHER)	<i>äah, äaht, ahr</i>	LXDATHER
'house' (vHOUSE)	<i>Huus, Kot</i>	W50
'into' (vINTO)	<i>(e)nenn, (e)nenna</i>	LXINTO
'knew' (vKNEW)	<i>wisst, wusst</i>	LXKNEW
'late' (vLATE)	<i>lot, spod</i>	W29
'little' (vLITTLE)	<i>kjleen, kjlien</i>	LXLITTLE
'moon' (vMOON)	<i>Mon, Mond</i>	W31
'often' (vOFTEN)	<i>foake(n), oft</i>	LXOFTEN
'out' (vOUT)	<i>(e)ruut, (e)ruuta</i>	LXOUT
'roads' (vROADS)	<i>Roode, Wäaj, Wäaje</i>	W04
'rub' (vRUB)	<i>riewe(n), rubb(l)e(n), schobb(r)e(n), strikje(n)</i>	LXRUB
'say (inf.)' (vSAY)	<i>saije(n), saje(n)</i>	LXSAY
'some' (vSOME)	<i>atlije/etliche, een Poa/een weenig, etwa(u)s/etwaut, some, walkje, waut</i>	LXSOME
'store' (vSTORE)	<i>Laufkje, Lode(n), Stua</i>	LXSTORE
'them (dat.)' (vDATTHEM)	<i>ähnt, ahn</i>	LXDATTHEM
'today' (vTODAY)	<i>von(d)oag, von(d)oagen</i>	LXTODAY
'tomato' (vTOMATO)	<i>Bockelzha(u)n/Bocklezha(u)n/Bottel- zhaun, Pomador, Temeeta/Temeetes, Tomato</i>	W55
'uncles' (vUNCLES)	<i>Onkels, Oohms</i>	LXUNCLES
'until' (vUNTIL)	<i>bat, bott</i>	LXUNTIL
'watermelon' (vWATERMELON)	<i>Arbuus, (H)erbuus, Me(r)loon, Rebuus, Wotameloon</i>	LXWATERMELON
'whether' (vWHETHER)	<i>auf, aus, es</i>	LXWHETHER
'would (2s.)' (vWOULD2S)	<i>wuddst, wu(r)scht</i>	LXWOULD2S

Table 9. Lexical variables and corresponding items in the *Fibel* Corpus.

Although variables such as these concentrate on differences at the level of individual words, it is also relevant to consider lexical variation that exists in larger collocations. In this respect, this study takes direction from the considerable body of research in corpus linguistics that explores linguistic patterning in the associations between words in their typical contexts of use (e.g., Stubbs 2001), as well as more recent work in both constructionist and usage-based approaches in linguistics (e.g., Bybee 2006) that argues that such associations play a significant role in the representation and processing of language. Although corpus-based studies of variation in lesser-

documented languages such as Plautdietsch have often struggled to employ the same analytical techniques as studies of other languages, in part due to the smaller corpora typically available for underdocumented languages and thus more restricted sizes of samples on which statistical measures of association and dispersion may draw (cf. McEnery & Ostler 2000, a.o.), this itself does not diminish the potential relevance of patterning between words and their surrounding environments. Although it is not easily possible to derive such associations reliably from existing Plautdietsch corpora, several multi-word units or lexically specified constructions can be proposed that demonstrate variation between speech communities. Several such items, assembled here under the label of ‘lexical-phrasal’ variables, are presented in Table 10 below.

<i>Variable</i>	<i>Variants</i>	<i>Source Items</i>
‘at the beginning’ (vATTHEBEGINNING)	<i>aum/aun Au(n)fang, em Au(n)fang, toom Aunfang(en), aun Bejinn/en de Aunbejinn</i>	CXATTHEBEGINNING
‘every year’ (vEVERYYEAR)	<i>aula Joah, jieda/jiedet Joah</i>	CXEVERYYEAR
‘in the evening’ (vINTHEEVENING)	<i>{em, en däm/dän/de, en’e} Owend, jäajen Owend, opp(’en) Owend, zeowe(n)s(t)</i>	CXINTHEEVENING
‘into the house’ (vINTOTHEHOUSE)	<i>em/en däm Huus, em Huus ‘enen, en daut Huus, en daut Huus ‘enen</i>	CXINTOTHEHOUSE
‘off of the wagon’ (vOFFOFTHEWAGON)	<i>{vom, von däm, von dän} Woage(n), {vom, von däm, von dän} Woage(n) ‘erauf</i>	CXOFFOFTHEWAGON
‘that (indef. relative clause)’ (vINDEFRELCLAUSE)	<i>daut, waut</i>	CXINDEFRELCLAUSE
‘that (neut. relative clause)’ (vNEUTERRELCLAUSE)	<i>daut, waut, woont</i>	CXNEUTERRELCLAUSE
‘that (pl. relative clause)’ (vPLURALRELCLAUSE)	<i>daut, daut doa, dee, waut, waut doa, woont</i>	CXPLURALRELCLAUSE
‘that day’ (vTHATDAY)	<i>dee Dag, däm Dag, dän Dag</i>	CXTHATDAY
‘the one (focus)’ (vTHEONE)	<i>däm/dän {eenzja/en, janja}, dee {eena, eenja, eenzja, janja, jansja, jeena, jeenja}</i>	CXTHEONE
‘without’ (vWITHOUT)	<i>ohne(n), met ohne(n)</i>	LXWITHOUT

Table 10. Lexical-phrasal variables and corresponding items in the *Fibel* Corpus.

It is apparent in the preceding table that whatever boundaries exist between strictly ‘lexical’ and more broadly phrasal or constructional items are often blurred, and thus open to question

concerning their classification into traditional linguistic levels of analysis. Several such lexical-phrasal items could alternatively be classified as morphological in nature: items such as ‘that day’ (CX<sub>THATDAY</sub>) or ‘the one’ (CX<sub>THEONE</sub>) concern lexical collocations that also involve the assignment of case, which could be viewed as a morphosyntactic, rather than a lexical-phrasal phenomenon. Likewise, it is possible to treat the pronouns used to introduce various kinds of relative clauses either as a matter of lexical choice or as being particular to certain larger phrases or constructions. Several of the constructions included in Table 10 could also be seen as instances of word choice in semi-fixed collocations, rather than as instantiations of a more generally productive morphological pattern. Treating such items at the outset as fully lexically instantiated constructions allows comparisons to be drawn between their morphological characteristics and those of other, more schematic constructions (which may differ in their regularity and productivity of case assignment, for instance), while still allowing for comparisons of word choice in the remaining, non-morphologically variable elements (e.g., *eenzja*, *eenja*, *janja*, *jansja*, *jeena*, or *jeenja* for ‘(that) one’ in CX<sub>THEONE</sub>). Here, the emphasis is less on the categorization such variation as belonging to one or another area of linguistic organization—there is little reason why such a classification must necessarily be all-or-nothing in nature, when some items may instantiate features of multiple categories and/or represent semi-productive or even wholly lexicalized patterns. Rather, the aim of this section is to ensure that such features do not escape attention, even if their overall productivity is limited and their assignment to particular linguistic categories remains open to further investigation.

This study makes a similar distinction between the lexical and lexical-phrasal variables seen above and the lexical-phonological variables summarized in Table 11 below. Unlike the phonological patterns in Section 4.2.1.3, which appear regularly in associated environments throughout the entire Plautdietsch lexicon, the patterns of variation treated here are restricted to small sets of lexical items and are not found elsewhere in the language. Thus, one may note that several modal verbs in Plautdietsch vary in their degree of reduction in second person singular forms (e.g., *kaunst* vs. *kau’st* ‘(you sg.) can’; LX<sub>CAN2S</sub>), or in the quality of their stem vowel in plural and infinitival forms (e.g., /e/ in *kjänne(n)*, *sälle(n)* vs. /ɛ/ in *kjenne(n)*, *selle(n)* ‘(you pl.) can, shall’; LX<sub>CANPL</sub>, LX<sub>SHALLPL</sub>). These patterns do not extend to the remainder of lexicon: it is not the case that nasals are consistently omitted before /s/, as in *kaunst* vs. *kau’st*, or that the

phoneme /e/ in one variety corresponds to /ɛ/ in another, as with *kjänne(n)* vs. *kjenne(n)*. Even so, such patterns may serve to distinguish varieties of Plautdietsch from one another and thus merit consideration here.<sup>32</sup>

<i>Variable</i>	<i>Variants</i>	<i>Source Items</i>
Reduction in ‘can (2s.)’ (vCAN2SREDUCTION)	FULL ( <i>kaunst</i> ), REDUCED ( <i>kau’st</i> )	LXCAN2S
Reduction in ‘can you (pl.)’ (vCANYOUPREDUCTION)	FULL ( <i>kjänn’ / kjenn’ (ji)</i> ), REDUCED ( <i>kjä’ / kje’ (ji)</i> )	LXCANYOUPL
Reduction in ‘could (2s.)’ (vCOULD2SREDUCTION)	FULL ( <i>kunnst</i> ), REDUCED ( <i>ku’st</i> )	LXCOULD2S
Reduction in ‘should (2s.)’ (vSHOULD2SREDUCTION)	FULL ( <i>sullst</i> ), REDUCED ( <i>su’st</i> )	LXSHOULD2S
Reduction in ‘supper’ (vSUPPERREDUCTION)	FULL ([ <i>ovən(t)kəs(t)</i> ]), REDUCED ([ <i>oŋkəst</i> ], [ <i>oŋkəs</i> ])	LXSUPPER
Realization of <i>au~ee</i> in ‘ate (pl.)’ (vAUeeATE)	AU ( <i>aute(n)</i> ), EE ( <i>eete(n)</i> )	LXATE
Realization of <i>au~ee</i> in ‘gave (pl.)’ (vAUeeGAVE)	AU ( <i>gauwe(n)</i> ), EE ( <i>jeewe(n)</i> )	LXGAVE
Realization of <i>au~ee</i> in ‘sat (sg.)’ (vAUeeSAT)	AU ( <i>saut</i> ), EE ( <i>seet</i> )	LXSAT
Realization of <i>au~ee~oo</i> in ‘took (pl.)’ (vAUeeOoTOOK)	AU ( <i>nauhme(n)</i> ), EE ( <i>neehme(n)</i> ), OO ( <i>noohme(n)</i> , <i>noohmpe(n)</i> )	LXTOOK
Realization of <i>au~auw~eiw</i> in ‘blue’ (vAUauwEiwBLUE)	AU ( <i>blau</i> ), AUW ( <i>blauw</i> ), EIW ( <i>bleiw</i> )	LXBLUE
Realization of <i>au~auw~eiw</i> in ‘grey’ (vAUauwEiwGREY)	AU ( <i>grau</i> ), AUW ( <i>grauw</i> ), EIW ( <i>greiw</i> , <i>jreiw</i> )	LXGREY
Vowel in ‘and’ (vANDVOWEL)	FRONT ([ <i>æn</i> , <i>ɛn</i> , <i>ɪn</i> ]), CENTRAL ([ <i>ən</i> ]), BACK ([ <i>ʌn</i> , <i>ʊn</i> ])	LXAND
Vowel in ‘can (pl.)’ (vCANPLVOWEL)	Ä ( <i>kjänne(n)</i> ), E ( <i>kjenne(n)</i> )	LXCANPL
Vowel in ‘horses’ (vHORSESVOWEL)	ÄA ([ <i>p<sup>h</sup>eəd</i> ]), EEA ([ <i>p<sup>h</sup>əɪəd</i> ]), IA ([ <i>p<sup>h</sup>iəd</i> ])	LXHORSES

32 The boundary between what is lexical choice and what is lexical-phonological variation may also be unclear at times. Items such as the lexical variables LXUNTIL and LX SAY illustrate this problem well: since the difference between *bat* /bat/ and *bott* /bɔt/ ‘until’ and *saje(n)* [sæjə(n)]~[sejə(n)] vs. *saijen* [sajə(n)] ‘say (inf.)’ is only a single vowel, these could be treated either as lexical-phonological pattern restricted to a single lexical item, or simply as a matter of lexical choice. As there is little evidence that either of these phonological patterns appear elsewhere in the lexicon, items such as these are treated here as lexical variables.

<i>Variable</i>	<i>Variants</i>	<i>Source Items</i>
Vowel in ‘shall (pl.)’ (vSHALLPLVOWEL)	Ä ( <i>sälle(n)</i> ), E ( <i>selle(n)</i> )	LXSHALLPL
Vowel in ‘was’ (vWASVOWEL)	ÄA ([vea]), EEA ([vøia]), IA ([via])	LXWAS

*Table 11.* Lexical-phonological variables and corresponding items in the *Fibel* Corpus.

Even while lexical-phonological variables are restricted to small sets of lexical items, it is still sometimes possible to distinguish larger patterns that hold common between such sets. Thus, one might observe the parallelism between stem vowels in modal verbs ‘can’ (*kjenne(n)* vs. *kjänne(n)*; LXCANPL) and ‘shall’ (*selle(n)* vs. *sälle(n)*; LXSHALLPL), or of the stem vowels in ‘ate’ (*aute(n)* vs. *eete(n)*; LXA TE) and ‘gave’ (*gauwe(n)* vs. *jeewe(n)*; LXGAVE). Although it is commonly assumed in the literature on Plautdietsch that speakers pattern coherently with respect to their selection of these variables (e.g., a speaker that uses the *ee* variant in ‘ate’ will also use the same variant in ‘gave’ and elsewhere), this is clearly not a logical necessity. In order to explore the possibility of larger, coherent patterns in these purportedly related variables, individual lexical-phonological variables are pooled to form the lexical-phonological macro-variables given in Table 12 below. As with similar definitions in preceding sections, these macro-variables allows larger patterns of variation to be explored without sacrificing attention to the individual features on which they depend.



<i>Variable</i>	<i>Variants</i>	<i>Sources</i>
Front vowels in pl. modals (vMODALPLVOWEL)	Ä ( <i>kjänne(n), sälle(n)</i> ), E ( <i>kjenne(n), selle(n)</i> )	vCANPLVOWEL, vSHALLPLVOWEL
Realization of <i>au~auw~eiu</i> (vAUAUWEIU)	AU ( <i>blau, grau</i> ), AUW ( <i>blauw, grauw</i> ), EIW ( <i>bleiw, greiw, jreiw</i> )	vAUAUWEIWBUE, vAUAUWEIWGUEY
Realization of <i>au~ee</i> (vAUÉE)	AU ( <i>aute(n), gauwe(n), kaum</i> ), EE ( <i>eete(n), jeewe(n), kjeem</i> )	vAUÉEATE, vAUÉEAVE
Realization of <i>ea</i> <sub>horses,was</sub> (vEAHORSESWAS)	ÄA ([p <sup>h</sup> eəd, vea]), EEA ([p <sup>h</sup> əiəd, vəia]), IA ([p <sup>h</sup> iəd, via])	vHORSESVOWEL, vWASVOWEL
Reduction in 2s. modals (vMODAL2SRREDUCTION)	FULL ( <i>kaunst, kunnst, sullst</i> ), REDUCED ( <i>kau'st, ku'st, su'st</i> )	vCAN2SRREDUCTION, vCOULD2SRREDUCTION, vSHOULD2SRREDUCTION

Table 12. Macro-variables and corresponding items in the *Fibel* Corpus representing recurring patterns across multiple lexical-phonological variables.

Several of these macro-variables merit further discussion. It should be noted that the alternation between *au~ee~oo* in ‘took (pl.)’ (vAUÉEooTook) has not been included in the *au~ee* macro-variable (vAUÉE). This is intended to allow the former pattern to be compared against the selection of variants for the latter, rather than grouping the two phenomena together prematurely. A similar approach has been taken with ‘sat (sg.)’ (vAUÉEsat), which has not been grouped together with other *au~ee* variables. The *ee* variant of this form has not been reported in studies of Mennonite Plautdietsch to date, although it is attested in the northern Polish dialect landscape. If this item invariably takes the form *au* for speakers of Mennonite Plautdietsch, then its inclusion in the *au~ee* macro-variable could give a false impression of the prevalence of variation in these vowels. Finally, the realization of *ea* across varieties of Plautdietsch is complicated, due in part to the history of convergence and divergence in associated lexical items in Middle Low German (cf. Lasch 1914, Siemens 2012: 80, 86). As a result, the corresponding macro-variable is defined here as having three possible contrasts to follow the distribution of all variants attested in Plautdietsch varieties, even though most varieties are likely to have only one or two of these forms.

#### 4.2.1.3 Phonological variables

While variation in the above lexical-phonological variables is most often instantiated by

only a few lexical items, it is also possible to identify larger, more broadly applicable patterns that recur over larger sections of the lexicon. Typically, such patterns appear in predictable phonological environments: thus, for some varieties of Plautdietsch, it is reported that word-initial /g/ appearing before a back vowel may be realized as a voiced velar fricative [ɣ], regardless of the lexical item (cf. Jedig 1966: 40, Moelleken 1967: 244, Moelleken 1972: 35). These more consistently instantiated patterns in phonemic realization are referred to here as phonological variables. Several such items are summarized in Table 13 below.

<i>Variable</i>	<i>Variants</i>	<i>Source Items</i>
Realization of coda /n/ before fricatives (vCODANPREFRICATIVE)	PRESENT ( <i>Aunfang</i> ), ABSENT ( <i>Au'fang</i> )	CXATTHEBEGINNING, LXSIXTYONE
Realization of /ea/ (vREALIZATIONEA) (*)	DIPHTHONG ([ia], [ɛa]), MONOPHTHONG ([e])	LXNEUTINDEFSoft, W12
Realization of initial /g/ before back vowels (vREALIZATIONGPREBACKVOWEL) (*)	STOP ([g]), FRICATIVE ([ɣ])	W18
Realization of final /g/ (vREALIZATIONFINALG)	STOP ([g]), FRICATIVE ([ɣ]), ABSENT	LXDAYS, LXTODAY
Realization of /ɔa/ before velars (vREALIZATIONOAPREVELAR) (*)	BACK ([oː], [oa]), FRONT ([øa], [ea])	LXCOOK, LXCOOKED, LXDAYS, LXMAKE, LXOFTEN, LXTODAY, W19, W39
Realization of /oa/ before velars (vREALIZATIONUAPREVELAR) (*)	BACK ([oː], [uː], [ua]), FRONT ([ya], [ɣa])	LXSUGARCOOKIES, W52
Realization of /r/ in complex codas (vREALIZATIONRCOMPLEXCODA)	ALVEOLAR (/r/, /r/), RETROFLEX (/ɻ/)	LXBECAME, LXBECAMESG, LXCORN, LXFARMER
Realization of final /rə(n)/ (vREALIZATIONFINALREN)	METATHESIS (/ərn/), NO METATHESIS (/rə(n)/)	LXBERRIES, LXPARENTS
Realization of final /st/ (vREALIZATIONFINALST)	CLUSTER ( <i>Owendkost</i> ), FRICATIVE ( <i>Owendkos'</i> )	LXBEEN, LXCAN2S, LXCOULD2S, LXSHOULD2S, LXSUPPER, LXWOULD2S
Realization of /fʃ/ in onsets (vREALIZATIONONSETTS) (*)	AFFRICATE ([fʃ]), FRICATIVE ([ʃ])	LXSIXTYONE, LXSUGARCOOKIES, W47
Realization of /u/ (vREALIZATIONU) (*)	BACK ([u]) FRONT ([y])	LXOUT, W50

Table 13. Phonological variables and their corresponding items in the *Fibel* Corpus. Starred items coincide with other variables in the phonemic inventory.

Several of these phonological patterns overlap with the coverage of the phonemic inventory introduced in Section 4.2.1.1, and thus provide greater coverage of possible intra-speaker variation in this feature than would a single item in the phonemic inventory. In all other cases, these variables refer to specific phonological environments and make clear predictions as to the realization of the phonemes under consideration. Barring any instances of phonetic indeterminacy that prevent specific realizations from being conclusively identified (cf. §4.2.3.4), the inclusion of such features in the present study is relatively straightforward.

#### 4.2.1.4 Morphological variables

Another important class of variables includes phenomena associated with the realization of word structure and inflection. These features are grouped together here as morphological variables, and are presented in Table 14.<sup>33</sup>

<i>Variable</i>	<i>Variants</i>	<i>Items</i>
Weak attributive adjective ending (ACC.M.SG.) (vENDINGADJACCMSG)	<i>-a (groota, bruuna)</i> <i>-en (grooten, bruunen)</i>	LXMASCACCDEFBIG, LXMASCACCDEFBROWN
Strong attributive adjective ending (ACC.N.SG.) (vENDINGADJACCNSG)	<i>-e (weakje),</i> <i>-et (weakjet)</i>	LXNEUTINDEFBROWN
Case selection, motion into (vCASEMOTIONINTO)	DATIVE ( <i>em, en däm</i> ), NON-DATIVE ( <i>en daut</i> )	CXINTOTHEHOUSE
Definite article (ACC.M.SG.) (vTHEACCMSG)	<i>däm, dän, dee</i>	LXMASCACCTHE,
Definite article (DAT.M.SG.) (vTHEDATMSG)	<i>däm, dän</i>	LXMASCDATTHE, CXOFFOFTHEWAGON
Definite article (M.SG.) in focus constructions (vTHEFOCUSCXMSG)	<i>däm (däm janja),</i> <i>dän (dän janja),</i> <i>dee (dee janja)</i>	CXTHEONE
Definite article (M.SG.) in possessive constructions (vTHEPOSSCXMSG)	<i>däm (däm Foarma sien),</i> <i>dee (dee Foarma sien)</i>	CXMASCPOSSTHE
Definite article (M.SG.) in time constructions (vTHETIMECXMSG)	<i>däm (däm Dag),</i> <i>dän (dän Dag),</i> <i>dee (dee Dag)</i>	CXTHATDAY

<sup>33</sup> Other morphological variation has also been identified in previous studies of Plautdietsch (e.g., a small class of nouns occasionally claimed to maintain an historical strong declension, receiving inflectional marking in non-nominative cases; cf. Quiring 1928: 85). While it is not always possible to target these features explicitly in the present study, given the constraints outlined in Section 4.2.2, these additional instances of variation have been documented in Appendix C, with the aim of encouraging their investigation in future studies.

<i>Variable</i>	<i>Variants</i>	<i>Items</i>
Demonstrative pronoun, ‘them’ (DAT.) (VTHEMDEMPRONDAT)	<i>dän, dee</i>	LXDATTHEMDEF
Ending, 3PL verbal <i>-e(n)</i> (VEN3PL)	<i>-e (aute, kjenne), -en (auten, kjennen)</i>	LXATE, LXCANPL, LXGAVE, LXHAVEAUXPL, LXHAVELEXPL, LXHELPED, LXTOOK,
Ending, infinitival <i>-e(n)</i> (VENINF)	<i>-e (koake, foahre), -en (koaken, foahren)</i>	LXCOOK, LXDRIVEINF, LXHAVEINF, LXMAKE, LXSAY, LXSINGINF, W02
Ending, final <i>-e(n)</i> in ‘often’ (VENOFTEN)	<i>-e (foake, föake), -en (foaken, föaken)</i>	LXOFTEN
Ending, final <i>-e(n)</i> in ‘without’ (VENWITHOUT)	<i>-e (ohne, met ohne) -en (ohnen, met ohnen)</i>	LXWITHOUT
Ending, nominal plural <i>-e(n)</i> (VENNOUNPL)	<i>-e (Beete, Bäare), -en (Beeten, Bäaren)</i>	LXBERRIES, LXXGIRLS, LXPARENTS, LXSUGARCOOKIES, W15, W20, W28 W18, W39
Ending, singular nouns in <i>-e(n)</i> (VENNOUNSG)	<i>-e (Goade, Woage) -en (Goaden, Woagen)</i>	
Ending, strong verb past participle <i>-e(n)</i> (VENPASTPART)	<i>-e (jeworde), -en (jeworden)</i>	LXBECAME, LXBEEN, LXEATEN, LXGIVEN, LXPAINTED
Past form of <i>halpe(n)</i> ‘to help’ (VPASTHELP)	STRONG ( <i>holpe(n)</i> ), WEAK ( <i>halpde(n)</i> )	LXHELPED
Past participial form of <i>jäwe(n)</i> ‘to give’ (VPASTPARTGIVE)	STRONG ( <i>jejäwe(n)</i> ), WEAK ( <i>jejäwt</i> )	LXGIVEN
Past participial form of <i>foawen</i> ‘to paint’ (VPASTPARTPAINT)	STRONG ( <i>jeforwe(n)</i> ), WEAK ( <i>jefoawt</i> ), IRREGULAR ( <i>jeforwt</i> )	LXPAINTED
Past participial prefix <i>je-</i> in <i>-eare(n)</i> verbs (VPASTPARTPREFIXEAREN)	PRESENT ( <i>jespazeat</i> ), ABSENT ( <i>spazeat</i> )	LXVISITED
Past participial prefix <i>je-</i> in non- <i>eare(n)</i> verbs (VPASTPARTPREFIXNONEAREN)	PRESENT ( <i>jejäwt</i> ) ABSENT ( <i>‘jäwt</i> )	LXBECAME, LXBEEN, LXEATEN, LXGIVEN, LXPAINTED
Plural form, <i>-en</i> singular nouns (VPLURALENNOUNSG)	<i>-ens (Schinkjens), -es (Schinkjes)</i>	LXHAMS, LXWAGONS
Possessive adjective ending, ‘our’ (DAT.M.SG.) (VOURDATMSG)	<i>-e (onse), -em (onsem), -en (onsen), NONE (ons)</i>	LXMASCDATOUR

Table 14. Morphological and morphosyntactic variables and corresponding items in the *Fibel* Corpus.

Although not further subdivided here, these variables comprise several distinct features. A

number of these variables deal with inflectional morphology relating to case, whether for adjectives (which may, in some varieties, have distinct inflectional endings for attributive singular masculine and/or neuter forms appearing in non-nominative cases, as with  $v_{\text{ENDINGADJACCMSG}}$  and  $v_{\text{OURDATMSG}}$ ) or articles (in particular argument structure or motion-related constructions that may show non-nominative inflectional marking on singular masculine and neuter forms, as in  $v_{\text{THEACCMSG}}$  and  $v_{\text{THEDATMSG}}$ ). Other variables treat nominal morphology not related to case inflection, as with the plural forms of nouns having singular forms ending in *-en* (e.g., *Schinkjen* ‘ham’ becoming either *Schinkjens* or *Schinkjes* in the plural;  $v_{\text{PLURALENNOUNSG}}$ ) or the realization of regular *-en* plural marking as either *-e* or *-en* (e.g., *Kaut* ‘cat’ becoming either *Kaute* or *Kauten* ‘cats’;  $v_{\text{ENNOUNPL}}$ ). Other variation in verbal morphology is considered here, as well, including the formation of past participles for verbs ending in *-earen* (which may sometimes take a *je-* participial prefix in Plautdietsch, unlike in Standard German;  $v_{\text{PASTPARTPREFIXEAREN}}$ ) and other verb classes, as well (where *je-* participial prefixes are typical, but may occasionally be absent in some Plautdietsch varieties, as per Quiring 1928: 95;  $v_{\text{PASTPARTPREFIXNONEAREN}}$ ).

Other morphological variables target the distinction between strong, weak, and irregular verbs, which distinguish classes of verbs based on the formal properties of their inflectional marking for tense.<sup>34</sup> As a core feature of the inflectional system of Plautdietsch, distinctions between these classes of verbs and their deviations from other, closely related languages are well documented (Quiring 1928: 94–105, Epp 2000, Neufeld 2000b, Neufeld 2000c, Siemens 2012: 171–183). Several of these studies note cross-dialectal variation in the preterite and past participial forms of some verbs (e.g., *jäwe(n)* ‘to give’ and *läse(n)* ‘to read’, for which both strong and weak past participles are attested; cf. Siemens 2012: 176–177), with variation of this kind captured by variables such as  $v_{\text{PASTPARTGIVE}}$  and  $v_{\text{PASTHELP}}$ .

---

34 Strong verbs are a limited set of lexical items that mark tense through ablaut in stem vowels (e.g., English *sing*, *sang*, *(have) sung*, or the cognate Plautdietsch forms *sinj*, *sung*, *jesunge(n)*). Weak verbs constitute a much larger and productive class that indicate tense through a *-t/-d* suffix (e.g., English *bake*, *baked*, *(have) baked*, or the cognate Plautdietsch forms *back*, *backt*, *jebackt*). Irregular verbs are either suppletive (e.g., *gohne(n)* ‘to go’, with forms *goh* (alongside 2s *jeihst*, 3s *jeiht*), *jingj*, and *jegohne(n)*; Epp 2000: 187) or combine the formal characteristics of both strong and weak verb classes (e.g., *froage(n)* ‘to ask’, with forms *froag*, *fruag*, *jefroagt*, where ablaut marks the preterite form *fruag* ‘asked’, but not the past participle *jefroagt* ‘(have) asked’).

It is also possible to identify larger patterns that may be shared by several morphological variables, as with variables related to *-e(n)* endings. While sharing the same phonological form, these endings appear in multiple constructional contexts, in some cases associated with nominal features (e.g., ending certain singular nouns, such as *Woage(n)* ‘wagon’ and *Goade(n)* ‘garden’, or marking plurality on nouns that take the *-e(n)* plural ending, such as *Beete(n)* ‘beets’), verbal morphology (e.g., strong past participle endings, as in *jeworde(n)* ‘(have) become’, infinitives such as *koake(n)* ‘to cook’, and plural subject inflectional endings, as in *se habe(n)* ‘they have’), and sometimes occurring in other lexemes (e.g., *foake(n)* ‘often’, *ohne(n)* ‘without’). As little is presently known about the distribution of variation across these constructional contexts—whether or not all of these items pattern similarly, or if individual features show different realizations of *-e(n)*—each of these variables is treated separately above. To consider the possibility of larger, coherent patterns shared by these distinct constructional contexts, however, several of these variables are also grouped into the macro-variables listed in Table 15.

<i>Variable</i>	<i>Variants</i>	<i>Items</i>
Ending, nominal <i>-e(n)</i>	<i>-e</i> ( <i>Woage, Beete</i> ), <i>-en</i> ( <i>Woagen, Beeten</i> )	vENNOUNSG, vENNOUNPL
Ending, verbal <i>-e(n)</i>	<i>-e</i> ( <i>jeworde, koake</i> ), <i>-en</i> ( <i>jeworden, koaken</i> )	vEN3PL, vENINF, vENPASTPART
Ending, other <i>-e(n)</i>	<i>-e</i> ( <i>foake, ohne</i> ), <i>-en</i> ( <i>foaken, ohnen</i> )	vENOFTEN, vENWITHOUT
Ending, <i>-e(n)</i>	<i>-e</i> ( <i>Woage, koake, foake</i> ), <i>-en</i> ( <i>Woagen, koaken, foaken</i> )	vENNOMINAL, vENOTHER, vENVERBAL

Table 15. Macro-variables and corresponding items in the *Fibel* Corpus representing recurring patterns across multiple morphological variables.

As with the lexical variables seen earlier, it is sometimes possible to argue for the assignment of certain morphological variables to other linguistic categories. While the forms of individual case inflection markers can be represented as morphological variables in their own right, it is clear that these forms are also dependent on the participation of other lexical items in larger constructional contexts (e.g., ones related to argument structure, focus, or location). As such, these constructions themselves might be suggested to present a more appropriate focus of investigation than the forms taken by individual inflectional morphemes, thus shifting the weight of analysis to morphosyntactic rather than more narrowly morphological features. While this

coding scheme treats these variables as morphological features, their contents are ultimately derived from corpus data that preserve the larger constructional contexts in which they occur. Thus, this approach does not preclude either alternate analyses of these phenomena at other levels of linguistic organization or further refinement (or even elimination) of these variables at later stages of analysis, should their present treatment prove insufficient for gaining insight into the patterning of associated variation.

#### *4.2.1.5 Syntactic variables*

As the preceding section noted, variation in words and word-forms may sometimes intersect with variation in the larger constructional contexts in which such items occur. In this study, differences in the configuration of concrete, phonologically specified items in more schematic and abstract frames will be treated as syntactic variation. The coding schemes adopted for this more schematic variation and the approaches taken to its assessment necessarily differ from many of the items reviewed in preceding sections, and therefore receive more attention below.

Syntactic variation differs from lexical, phonological, and morphological variation in the lesser degree of attention it has generally received in studies of Plautdietsch. As Section 1.2.2 observed, this emphasis on linguistic phenomena below the level of the utterance is a trait shared with traditional European dialectology, where the earliest systematic surveys of Romance and Germanic dialects focused largely on the elicitation and careful phonetic transcription of lexical and morphophonological differences thought to best characterize geographically correlated variation. The resulting uneven representation of phenomena at other levels of linguistic organization is in part a consequence of the theoretical influences that acted on the formation of these traditions, perhaps most prominently the early comparative linguistic tradition, in which evidence from dialect geography featured prominently in debates over the regularity and diffusion of sound changes. No doubt also implicated in this persistent bias were the tools at the disposal of earlier generations of researchers, for whom the collection of realistic samples of spoken language and natural discourse was practically impossible before the advent of technologies for recording and reproducing connected speech, which in turn enabled the analysis of spoken language and the patterns observed within it (cf. Newman 2008).

This general dearth of attention to utterance-level phenomena in dialectological and dialectometric research persisted until relatively recently, with the development of contemporary dialect atlases for languages such as Dutch (Barbiers et al. 2005) and Catalan (Prieto, Cabré & Vanrell 2010), which provided substantial and emphatic representation of prosodic and syntactic variation correlated with geography. In other, related research traditions, too, variable syntactic phenomena have only recently begun to assume the same prominence as morphophonological and lexical differences, although for different reasons. In the case of variationist sociolinguistics, issues of data sparseness in smaller sociolinguistic corpora that lessened the amenability of much reported syntactic variation to quantitative statistical analysis, coupled with questions of the overall applicability of the notion of the sociolinguistic variable to syntactic variation (cf. Milroy 1987, ch. 7), frequently limited investigation in these areas to languages for which large-scale corpora were available (Moisl 2009).

Whatever its historical roots, this overall pattern of neglect for investigations of syntactic phenomena has held for studies of Low German, as well. As has long been noted in the literature on this language group, syntactic phenomena present significant gaps in accounts of varieties of Low German, receiving considerably less attention than aspects of their phonology, morphology, and lexical choice (Meier 1978: 290, Saltveit 1983: 282, Cox 2008: 14–15). Plautdietsch presents no exception to this trend: although extensive studies exist of the Plautdietsch lexicon (e.g., Thiessen 1963), phonology (e.g., Jedig 1966) and morphology (e.g., Buchheit 1978), even full-length grammatical descriptions of Plautdietsch varieties give syntactic phenomena relatively little attention. With the exception of Jedig (1969) and a series of recent studies by Kaufmann (2003b, 2005, 2007, 2008) and Cox (2008, 2011a), few resources exist on constructions above the level of the word in Plautdietsch, and fewer still that consider variation in this area.

The order of verbal elements in largely schematic clausal constructions presents one area in which syntactic variation has been reported in investigations of Plautdietsch, namely in so-called verb clusters (Bech 1955, Evers 1975, Wurmbrand 2004). In Plautdietsch as in other continental West Germanic languages, the unmarked order of constituents differs between declarative main clauses and dependent clauses. In most declarative main clauses, an inflected verb is found in second (Wackernagel's) position, as in (1). In dependent clauses, the inflected



verb typically appears closer to the end of the clause, as in (2):<sup>35</sup>

- (1) T’Huus **säd**’ wi “Bockelzhonn.”  
 at.home said we tomato  
 ‘We **said** “*Bockelzhonn*” at home.’ (comment on lexical item for ‘tomato’)  
 (M08, 2011-09-13 (02), 28m36s405–28m37s935)
- (2) Ahn deed daut leet, [ daut se soo vāl Kjoaschen **jejäten**<sub>2</sub> **hauden**<sub>1</sub>. ]  
 them did that sorry [ that they so many cherries eaten had.PL ]  
 ‘They were sorry that they **had**<sub>1</sub> **eaten**<sub>2</sub> so many cherries.’  
 (M08, 2011-09-13 (02), 13m39s610–13m43s160)

Verb clusters arise where these clause type-dependent orders intersect with verbal complementation constructions. In continental West Germanic languages like Plautdietsch, complex predicates are often formed by verbs introducing other verbs as their complements. In Plautdietsch, these complements may be bare infinitives (3a), *too*-marked infinitives (3b), or past participles (3c) (Bech 1955, 1957). In the following examples, subscript numerals give the order of complementation, with verb  $v_i$  introducing verb  $v_{i+1}$  as its complement:

- Bare infinitive complement*
- (3) a. Ekj si emma dee jansja, wäm doa **mott**<sub>1</sub> **weeden**<sub>2</sub> em Goaden.  
 I am always the one whom there must weed.<sub>INF</sub> in.the garden  
 ‘I’m always the one that **has**<sub>1</sub> to **weed**<sub>2</sub> in the garden.’  
 (F17, 2011-10-27 (01), 15m58s815–16m02s285)
- Too-marked infinitive complement*
- b. He **proowt**<sub>1</sub> ‘nen **too komen**<sub>2</sub>.  
 he tries in to come.<sub>INF</sub>  
 ‘He’s trying to come in.’  
 (M15, 2011-08-04 (01), 13m39s345–13m40s755)
- Past participle complement*
- c. Dit sen’ dee Wäaj, waut ons Voda **haft**<sub>1</sub> **jemöakt**<sub>2</sub>.  
 these are the roads what our father has built.<sub>PART</sub>  
 ‘These our the roads that our father **has**<sub>1</sub> **built**<sub>2</sub>.’  
 (F17, 2011-10-27 (01), 03m16s035–03m19s395)

As these examples suggest, verbal complementation constructions are commonly found in modal

<sup>35</sup> Wherever possible, examples drawn from the Fibel Corpus described in Section 4.2.3 are cited with full reference to the contributor(s) (e.g., F17, M15; see Table 18), the corresponding corpus source (e.g., 2011-10-07 (01), referring to the first part of the resource created on 2011-10-27; see Table 17), and the starting and ending times in this recording in which this example occurs (e.g., with the interval 15m58s815–16m02s285 beginning nearly 16 minutes into the recording and ending shortly after the 16 minute mark, with times given in minutes, seconds, and milliseconds).

constructions (i.e., where a modal verb introduces a bare infinitive, abbreviated MOD-INF), passives (i.e., where a form of the auxiliary verb *woare(n)* ‘to become’ introduces a past participle, abbreviated AUX<sub>PASS</sub>-PART), and in perfect tense constructions (i.e., where a form of the auxiliary verbs *habe(n)* ‘to have’ or *senne(n)* ‘to be’ introduces a past participle, abbreviated AUX<sub>PERF</sub>-PART). Verbal complementation includes not only these high-frequency two-verb constructions, but also includes less common cases in which a complement verb introduces another verbal complement of its own. These include instances like (4), where an auxiliary verb introduces a modal verb, which in turn introduces a bare-infinitival complement (abbreviated AUX-MOD-INF):

- (4) Dise Migjen wissten nich, daut dee **hauden<sub>1</sub> sultt<sub>2</sub>** buten [.] **bliewen<sub>3</sub>**.  
 these mosquitos knew not that they had should outside stay. INF  
 ‘These mosquitos didn’t know that they **should<sub>2</sub> have<sub>1</sub> stayed<sub>3</sub>** outside.’  
 (F29, 2011-08-04 (01), 15m27s755–15m30s885)

These constructions present a well-documented point of variation among West Germanic languages, where the range of acceptable orders of verbs varies both across varieties and between constructions (Wurmbrand 2004, Sapp 2006, Dubenion-Smith 2010). Thus, in a modal-infinitival (MOD-INF) construction appearing in a verb-final clause, the order  $v_1-v_2$  is not acceptable in Standard German, as seen in example (5a). By contrast,  $v_1-v_2$  is the only licit order for the same construction in West Flemish (5b), while Standard Dutch in (5c) permits both alternatives (examples from Cox 2011a: 551):

- (5) a. ...*weil er das Buch \*muss<sub>1</sub> kaufen<sub>2</sub> / kaufen<sub>2</sub> muss<sub>1</sub>*.  
 ‘because he had<sub>1</sub> to buy<sub>2</sub> the book’ (Standard German; Wurmbrand 2004: 74)
- b. ...*da Valère dienen boek wilt<sub>1</sub> kuopen<sub>2</sub> / \*kuopen<sub>2</sub> wilt<sub>1</sub>*.  
 ‘because Valère wants<sub>1</sub> to buy<sub>2</sub> this book’ (West Flemish; Wurmbrand 2006: 331)
- c. ...*dat Jan het boek kan<sub>1</sub> lezen<sub>2</sub> / lezen<sub>2</sub> kan<sub>1</sub>*.  
 ‘that Jan can read the book’ (Standard Dutch; Wurmbrand 2006: 324)

Similar differences in acceptability are encountered in Mennonite Plautdietsch, with variation attested among speech communities both internationally and in Saskatchewan specifically (cf. Kaufmann 2005; Cox 2008, 2011a). Given the frequency of these constructions and previous reports of noteworthy variation among Mennonite Plautdietsch speech communities in their use, these verb cluster constructions are taken here as a point of departure for the investigation of

syntactic variation in Plautdietsch.

As previous studies of cognate constructions have noted (e.g., Kaufmann 2005, Dubenion-Smith 2010), it is important to consider several contextual parameters relevant to verb order in these constructions. Different construction types may show different categorical distinctions or gradient preferences with respect to verbal order: a two-verb perfective construction may not show the same distribution of constituent orders as a two-verb modal construction, for example, despite both having the same number of verbal constituents. For the present study, four of the most frequent complementation constructions reported in the corpus-based study of Cox (2008) are adopted as variables, namely modal-infinitival (MOD-INF), perfective (AUX<sub>PERF</sub>-PART), passive (AUX<sub>PASS</sub>-PART), and auxiliary-modal-infinitival (AUX-MOD-INF) constructions. As common two and three-verb constructions, this selection permits these robustly attested constructions to be targeted for comparison, without introducing additional complexity (with more verbal complements) or uncommon constructional choices (for less frequent complementation constructions).

In addition to controlling for specific complementation constructions, it is also necessary to consider differences related to clause type and transitivity. Dependent clauses in which complementation constructions occur may differ in their function, serving variously as causal (CAUS), complement (COMP), conditional (COND), or relative (REL) clauses. At the same time, the final complement verb may either be intransitive (INTR) or introduce a direct or oblique object (DO-IO), contributing to differences in syntactic weight shown to be relevant to alternations in Plautdietsch verbal complementation constructions (Cox 2008: 169, Cox 2011a). As summarized in (6), considering all logical combinations of these three factors—complementation construction, clause type, and transitivity—allows for balanced comparison of the effects of these important predictors. The subsequent treatment of these factors as syntactic variables in all 32 logically possible combinations is discussed further in Section 4.2.2.2.

- (6)     { AUX-MOD-INF, AUX<sub>PASS</sub>-PART, AUX<sub>PERF</sub>-PART, MOD-INF } X     *Construction*  
          { CAUS, COMP, COND, REL } X                             *Clause type*  
          { INTR, DO-IO }   *Transitivity*

#### 4.2.1.6 Summary

The preceding sections have outlined a series of lexical, phonological, morphological, and syntactic features reported to vary among speakers of Mennonite Plautdietsch varieties. This inventory, while not exhaustive of variation in the language, is nevertheless extensive. In addition to providing coverage of 55 phonemic and allophonic contrasts, it includes some 44 lexical variables, 11 lexical-phrasal variables, 16 lexical-phonological variables (as well as five macro-variables derived from these features), 11 phonological variables, 23 morphological variables (as well as four derived macro-variables), and three classes of syntactic features related to clausal word order in 32 unique combinations, for a total of 108 primary features. The range of variants attested for all non-syntactic and non-phonemic variables, as well as their corresponding source items in the corpus described below, are summarized in Appendix C.

These variables present a starting point for an investigation of linguistic variation in Mennonite Plautdietsch in the Saskatchewan Valley. Having been drawn in large part from the literature on linguistic variation in Mennonite Plautdietsch-speaking communities, these features are likely to be shared with other groups, and thus facilitate comparisons not only within the Saskatchewan Valley, but also with communities elsewhere in the Russian Mennonite diaspora. Considering variables representing patterns at different levels of linguistic organization—from sounds, to words, to multi-word constructions—also has potential merit in avoiding too narrow an initial focus that may result in a false picture of variation in Saskatchewan Valley communities (cf. §1.2.1, §1.2.3). As will be seen in Section 5.2, this multivariate approach also allows for potentially differences in variation in different facets of linguistic organization to be investigated, something that would not be trivial to pursue with a more restricted set of variables.

Other aspects of this set of variables may merit attention in continued research. Although the inclusion of lexical-phrasal and syntactic variables here is arguably beneficial, it still does not fully counterbalance the heavier attention given to lexical, phonological, and morphological variation in this sample. This unevenness follows in large part from reliance on the existing literature on variation in Mennonite Plautdietsch, where differences in these areas are most robustly attested. A similar tendency in this literature to favour variation represented by discrete, symbolic linguistic units—whether phonemes, morphemes, lexemes, or other constructions—rather than sub-symbolic or continuous ones is also apparent. Again, this results from drawing

on the extant literature to identify points of variation, but brings attention to the need for forms of documentation that allow such questions of sub-symbolic variation to be investigated in the future, even if their exploration lies outside of the scope of this study. Although limitations such as these do not prevent an initial exploration of variation in Mennonite Plautdietsch, they do call for added scrutiny when interpreting the results of subsequent analyses that are unable to consider variation in these other important aspects of linguistic organization.

While the set of variables introduced in the preceding sections represents a considerable sample of the linguistic variation reported to exist in diasporic Mennonite Plautdietsch speech communities, the question remains of how to incorporate these observations into a study of Saskatchewan Valley. The problems of integrating these variables into planned language resource materials, observing and documenting instances of variation in this context without drawing undue attention to the corresponding variables' presence, and developing the resulting documentation into an internally consistent, annotated corpus accessible for research are therefore taken up in the following sections.

#### 4.2.2 Eene Plautdietsche Fibel

##### 4.2.2.1 Motivations

Having identified a range of features in Mennonite Plautdietsch varieties that are reported to vary between speech communities, the discussion now turns again to the means by which such variation is to be observed and studied. Section 4.2 laid out several desiderata for the development of community language resources in the Saskatchewan Valley. According to these criteria, not only should these resources be relevant and accessible to local language-related initiatives, but they should also aim not to presume or require contributors' proficiency with written forms of Plautdietsch. Discussions of possible outcomes that would satisfy these aims led to the proposal of a Plautdietsch-language *Fibel*, or primer, modelled after the High German *Fibel* of the traditional Russian Mennonite educational system (cf. §2.5). While this suggestion emerged in the context of local, language resource-related needs, it appears entirely possible for the *Fibel* to contribute to the linguistic aims of this study, as well, allowing for a large number of linguistic variables to be given consistent and systematic attention while guarding against sparseness of data. This section considers the strengths and limitations of a *Fibel* as a means of

investigating linguistic variation in the Saskatchewan Valley, before proceeding to describe how this primer and an associated corpus were developed in collaboration with contributors from these communities.

Given the importance of not prematurely limiting exploration of variation in these communities to only a few predictors, and given the lack of prior documentation that could provide complementary information on patterns of linguistic variation in Saskatchewan Valley communities, some form of standardized task common to all contributors would ultimately seem necessary to ensure consistent coverage of a wide range of linguistic variables. At least in the initial phases of research, such considerations motivate against approaches that centre on the observation of casual, vernacular speech without the use of a common task such as reviewing the *Fibel*. Although spontaneous speech no doubt presents considerably richer material from both historical-cultural and linguistic perspectives, for the purpose of an initial investigation of variation in these communities, it is unlikely that a significant number of the variables identified in the preceding sections would appear with sufficient frequency in samples of spontaneous discourse to permit systematic comparison across contributors and communities (or at least not without recording many more hours of speech than could feasibly be annotated using presently available technologies and resources). In practical terms, the use of spontaneous discourse as the primary source of linguistic evidence in this study would imply that only the most common variables (e.g., those phonemic, phonological, and morphological variables that appear in the most frequent constructions) can realistically be brought into analysis, leaving others that are more sparsely attested (e.g., lexical and syntactic variables) essentially on the cutting-room floor.

It is clear that the selection of a standardized survey instrument such as a *Fibel* does not necessarily preclude other forms of targeted investigation, or even some representation of spontaneous speech to complement these more controlled linguistic tasks. In the present case, however, such samples of unscripted discourse were not identified by members of the Saskatchewan Valley communities as being of pressing interest within the framework of this project. Additional recorded interviews or conversation sessions were not seen as being of as immediate need as other kinds of language resources, and even were suggested to act as a potential source of discouragement for contributors who were willing to assist in developing a Plautdietsch-language *Fibel*, but were much more hesitant about being recorded in casual

conversation.<sup>36</sup> Similarly, the possibility of introducing another kind of controlled linguistic task that could offer ample coverage of variables, such as reading a supplied passage of text, was not seen as appropriate to the local sociolinguistic situation. Whether educated in the traditional Mennonite educational system (for the very oldest generation, in Mennonite Standard German) or in the provincial educational system (for those born after approximately 1920, in English), few members of the Plautdietsch-speaking communities in the Saskatchewan Valley have received any formal education in their first language. Many proficient Plautdietsch speakers in the Saskatchewan Valley were thus reticent to attempt reading Plautdietsch-language materials out loud, given the relatively unfamiliar nature of the many orthographic systems currently in use.

Both the sociolinguistic situation in the Saskatchewan Valley and the lack of prior documentation for these communities present motivation for the development of a specialized language resource—one that might provide adequate representation of the variables of interest, be accessible for community language initiatives, and take into consideration the general preference on the part of local community members for their contributions to the development of this resource to be made in spoken, rather than written form. While these requirements place notable constraints on the structure of the *Fibel*, other aspects of the sociolinguistic situation in these communities afford somewhat greater latitude. In particular, it was noted in Section 2.6 that the present-day Saskatchewan Valley Mennonite Plautdietsch speech community is universally bilingual in English, with some speakers also being capable of reading and writing forms of Standard German. No monolingual adult speakers of Plautdietsch are reported in the area today. This high level of proficiency in both English and Mennonite Plautdietsch admits the possibility of employing a translation task—one into which a wide range of variables identified in the preceding sections might be incorporated into English-language prompts, thus avoiding overt attention being drawn to the variation of interest and limiting the degree to which the use of other varieties of Plautdietsch or Huagdietsch might lead to different responses.

While there are thus several reasons to consider developing documentary resources in this

---

36 As Section 1.2.1 observes, interviews or conversations conducted for research purposes in the language also risk introducing substantial bias through the selection of interlocutors: with no clear sense at the outset of the social markedness of differences that may exist between local varieties, it is difficult to anticipate what effect a particular Plautdietsch-speaking interlocutor (whether a researcher or other community partner) may have on contributors' vernacular speech patterns.

way, there are necessarily also drawbacks to such an approach. Perhaps most obviously, a corpus developed using a specialized translation task, although able to provide systematic coverage of a large number of embedded variables across a considerable number of speakers, is more limited in its potential uses than a corpus of spontaneous speech. Careful attention to associated metadata and recording methods may serve to produce a record more amenable to a range of future reuses, but does not fully address the limitations inherent in the underlying elicitation tasks. For one, a specialized corpus based on observations of these kinds of circumscribed linguistic acts can provide only scant information about the relative frequency of individual features. Certain common variables may be open to comparisons of their frequency across speakers, but in many cases, the constrained nature of the task and the repetition of the same prompts across speakers may limit the degree to which the frequency of occurrence of one variant over another can be accurately assessed.

Moreover, given the underdocumented state of Saskatchewan Valley Mennonite Plautdietsch, it may be difficult to gauge the extent to which the responses that form the basis of this kind of specialized corpus are typical of contributors' speech, or have been affected by the nature of the task. Although Kaufmann (2005) argues that translation is not at all atypical in pervasively multilingual Russian Mennonite communities, and that the almost daily nature of translation for enclaved linguistic minorities makes it reasonable to assume that the degree of interference introduced through translation tasks is less significant than it might otherwise be, what remains in question is the *extent* of such effects, rather than their occurrence at all. Some deviation from everyday speech norms might be expected to be introduced in this context, whether in atypical lexical or constructional choices (e.g., in an extreme case, translating word-for-word, leading to potential divergences from what would otherwise be typical for this speech community and for the language in general; or, in a more normal situation, being led to select constructions in the target language that bear some formal or functional similarity to the translation prompts) or in more careful, slower, and thus potentially prosodically divergent speech. Whereas one might suppose that overt focus on a linguistic task such as this may favour standard-like forms of speech, it should be recalled that the relationship between 'roofless' (*dachlos*) local varieties of Plautdietsch and other potential superordinate standard languages (e.g., Mennonite Standard German, English) or even other varieties of Plautdietsch is not known



at the outset, and thus the nature of any convergence toward one or another standard also remains open to debate.

One might seek to correct for such potential imbalances through comparison with other sources of information on local Plautdietsch varieties. Although a small, annotated corpus of written Saskatchewan Plautdietsch texts exists (cf. Cox 2008), being limited at present to the works of two Mennonite authors (only one of whom is from the Saskatchewan Valley proper), it is likely too limited to be of much assistance in situating the variation found in a more specialized corpus within the broader context of the Saskatchewan Valley—especially since the dialect affiliations of the authors in the corpus are themselves a matter for further research. Other written samples of Saskatchewan Valley Mennonite Plautdietsch are found in Brednich (1977, 1981), Friesen (1988), Guenter et al. (1995), and Driedger (2011), although these sources are no more extensive than the aforementioned corpus. As such, it is difficult to attempt to identify potential imbalances in this specialized corpus through reference to a more general one. Rather, one might hope that the sharper focus of the specialized corpus might be useful in analyzing these other, more naturalistic records, providing a clearer sense of their dialectal placement and pointing to areas where further documentation may be needed.

The decision to develop a specialized, translation task-based corpus is thus an imperfect solution to a difficult problem, albeit one that offers certain advantages in this situation. The ability of a carefully planned translation task to target multiple aspects of linguistic variation simultaneously across a large number of speakers in a tractable, systematic manner is of considerable importance, given the scope of this study and the absence of prior documentation. This same lack of comparative linguistic information, however, combined with the somewhat unusual nature of the linguistic task at hand, makes it difficult to assess the degree to which this choice of tasks results in atypical data. Certainly, this record cannot be the final word on linguistic variation in Saskatchewan Valley Mennonite Plautdietsch, nor offer the sole source of linguistic information for these communities; nor is this the role intended for the *Fibel* Corpus or the linguistic tasks behind it. Rather, the aim of this choice of methods is twofold, attempting to respond appropriately to community interests in the creation of this kind of language resource while furnishing the information necessary for a first foray into local patterns of linguistic variation, such that the questions raised and resource needs identified in the process might be

addressed more extensively in further work.

#### 4.2.2.2 *Design and review*

Given the decision to observe linguistic variation primarily through a collaboratively developed community language resource, the question arises of how such a resource is to be designed to meet the needs of both local communities and linguistic research. At the outset, it is important to ensure that the variables identified in the preceding sections are adequately represented, without drawing undue attention to their presence or rendering the resulting materials unwieldy for use in either a translation task or as a learner's resource. As well, since a translation task introduces the risk of variation between contributors in their interpretation of prompts, care should be taken to reduce the potential ambiguity of such items wherever possible. Both of these concerns point to the importance of the design of the *Fibel*, ensuring that these aims are met in the final version of the primer while keeping the associated translation task both manageable and relatively constrained.

The format of the *Fibel* presents opportunities to integrate many of the above-highlighted variables quite naturally. Much as the traditional Russian Mennonite *Fibel* proceeded through all of the sounds of Standard German in sequence, adopting a similar approach here allows for each of the phonemic categories identified in Section 4.2.1.1 to be integrated through the target words in Table 8. This provides coverage of all major phonemic contrasts in maximally consistent phonological environments without departing markedly from the basic cultural model for this kind of resource. This study therefore opted to begin with these phonemic categories, with each such sound presented on a separate page of the *Fibel*. Rather than display the Plautdietsch-language target word on the page (and thus potentially bias translations in favour of particular variants), its English equivalent was given, instead. Thus, for the phoneme /g/, the English word 'garden' was given on the page, rather than its Plautdietsch target *Goade(n)*, since the distinct variants *Goade* and *Goaden* are both possible. To constrain the range of possible translations and to keep with the traditional *Fibel* format, each page was also identified with the Plautdietsch sound found in the target item. Thus, for the page targeting the phoneme /t/, not only was the English word 'tame' presented, but also the grapheme <t> with which this sound is generally

associated in Plautdietsch.<sup>37</sup> This approach proved beneficial in many cases where the English term could have more than one translation into Plautdietsch: as one contributor commented in (7), both *mack* and *tohm* would be appropriate Plautdietsch translations of ‘tame’, albeit with slightly different meanings:

- (7) “Tohm” wudd senn', wan du daut tohm jemöakt haudst. Un “mack,” // daut wudd  
*tohm* would be if you it tame made had and *mack* that would  
 senn', [.] *like*, // um, // “*quiet*” or “*tranquil*.”  
 be like um quiet or tranquil  
 ‘*Tohm* would be if you had made it tame. And *mack*, that would be, like, um, “quiet” or  
 “tranquil.”’  
 (F17, 2011-10-27 (02), 08m13s980–08m21s195)

The presentation of an English-language prompt word and a Plautdietsch-language phoneme found in the translation was thus in line with the format of the traditional Russian Mennonite *Fibel*, provided full coverage of the Plautdietsch phonemic inventory, and militated against highly divergent translations.<sup>38</sup> With these example words having been selected, the remaining non-phonemic variables were integrated into example sentences for each of the target words, much as in the traditional Standard German *Fibel*. Wherever possible, variables were included in more than one example sentence in the *Fibel* to provide redundant (and, thus, ideally more

---

37 Although orthographic conventions vary between diasporic Plautdietsch-speaking communities, many Canadian Plautdietsch authors (with the exception of Loewen 1996, 1998, who advocates a phonemic orthography that more closely resembles English in some of its digraphs) endorse some variant of either the orthographies proposed by earlier standardization efforts (e.g., Reimer 1982, Epp 1996) or adopted by the recent Plautdietsch-language translation of the Bible. The general convergence of these proposals on common conventions for representing most phonemic categories (with the notable exception of <kj>, <tj>, or <tch> for /k~tj~c/) makes the presentation of such phonemes here less problematic than it might otherwise be; cf. §4.2.3.3.

38 The presentation of an orthographic phoneme target sound could be argued to potentially bias contributors in favour of certain variants that bear close resemblance to the corresponding sounds in a more familiar written language (e.g., presenting <öa> for pre-velar fronted /ɔa/ could lead some contributors to follow Standard German conventions and round or front this diphthong, even if this would not be their normal practice). This was given explicit attention in early pilots of the *Fibel* and in the actual study, and little or no such influence was noted. In general, contributors made reference to these phoneme targets only in cases where more than one Plautdietsch translation was possible (cf. *mack* vs. *tohm* for ‘tame’ above), and demonstrated the same selection of variants in these targeted items as they did in the same items appearing less prominently in example sentences elsewhere in the *Fibel*, suggesting little overall influence on vernacular preferences.

extensive and reliable) coverage. These example sentences were modelled after naturally occurring examples drawn from the Saskatchewan Mennonite Plautdietsch corpus described in Cox (2008). This comparison with unelicited examples of spontaneous language use provided counterbalance against potentially unconventional choices of sentence structures or lexical items in such sentences, offering somewhat more natural frames for the variables of interest.

Incorporating the syntactic variables from Section 4.2.1.5 further called for these sentences to reflect all possible combinations of clause type, transitivity, and verbal complementation construction, requiring additional planning to ensure that all such options were not only represented, but distributed across the *Fibel* in such a way as to minimize potential priming from earlier responses of the same construction or clause type (e.g., by including prompt sentences throughout that did not target any syntactic variables, and by ensuring that sentences of the same construction or clause type never occurred directly after one another).

The result of this initial design process was a preliminary set of 55 example words and sentences, intended to be presented in English for translation into Plautdietsch. In consultation with several members of the Saskatchewan Valley Mennonite community who served as local advisors on the development of the *Fibel* (see fn. 28, p. 68), this initial set of words and sentences was subjected to several rounds of piloting and revision, substantially improving the perceived ‘naturalness’ of the prompt sentences and lessening the ambiguity of particular items. Other suggestions targeted particular features of the prompt sentences that required adaptation to match conventional uses of English in Saskatchewan Mennonite communities. It was recommended that the term ‘lunch’ be replaced with ‘dinner’, for instance, which was seen as the more typical term of reference for the mid-day meal; and that ‘kids’ be replaced with ‘children’, as some members of the community reportedly considered the former term to be inappropriate.<sup>39</sup> Similarly, it was recommended that each item in the *Fibel* be accompanied by a picture or illustration that could serve as a visual prompt for the elements being translated. Not only would this improve the overall attractiveness of the final resource, it was suggested, but also provide further useful disambiguation for translation. Thus, in a prompt sentence containing the term ‘to

---

39 Similar advice is given by the Mennonite Central Committee of Ontario, which recommends that non-Mennonite community service workers not use the term ‘kids’ in interactions with Plautdietsch-speaking Mennonites to avoid possible offence; see <http://mcco.ca/lowgerman/relating/communication> (retrieved February 22, 2014).

smoke’, providing an accompanying picture that contained a pipe or cigarette might encourage speakers to use the term *schmeakje(n)* ‘to smoke (tobacco)’ rather than *reakjre(n)* ‘to smoke (let off smoke, cure meat)’. Using images in the public domain or made available under suitable Creative Commons licenses, each page of the *Fibel* was associated with an image that members of the advisory group considered appropriate.<sup>40</sup> The result of these revisions was the final set of prompt words, sentences, and syntactic frames used in the *Fibel*, given in Table 16 below.

---

<sup>40</sup> These images, their sources, and their presentation in the final version of the *Fibel* are reproduced separately at <http://hdl.handle.net/10402/era.39445>.

<i>ID</i>	<i>Sound</i>	<i>Word</i>	<i>Sentence</i>	<i>Clause</i>	<i>Construction</i>	<i>Transitivity</i>
S01	a	/a/ <i>Bad</i> /bad/ 'bed'	The little girls shouldn't jump on the bed.		(free-form)	
S02	ä	/e/ <i>bädt</i> /bet/ 'prays'	The old man is praying that his grandchildren have all slept well.	COMP	AUX <sub>PERF</sub> -PART	INTR
S03	äa	/εa/ <i>Bäa</i> /bea/ 'berry'	Children, can you pick her some strawberries and blueberries?		(free-form)	
S04	äaj	/εaɪ/ <i>Wäaj</i> /veaɪ/ 'roads'	These are the new roads that our father has made.	REL	AUX <sub>PERF</sub> -PART	DO-IO
S05	au	/aʊ/ <i>Bauss</i> /baʊs/ 'boss'	Our boss says that we can have dinner early today.		(free-form)	
S06	auj	/aʊɪ/ <i>Krauj</i> /kraʊj/ 'crow'	Did you know that a crow can sing nicely?	COMP	MOD-INF	INTR
S07	b	/b/ <i>Boa</i> /bʊa/ 'bear'	You should leave the big brown bear alone.		(free-form)	
S08	ch	/x/ <i>Dag</i> /dax/ 'day'	We hadn't wanted to go out that day because it was supposed to be hot.	CAUS	AUX-MOD-INF	DO-IO
S09	d	/d/ <i>Dola</i> /dola/ 'dollar'	All that they have is sixty-one dollars.		(free-form)	
S10	e	/ə/ <i>besied</i> /bəzid/ 'beside'	The cat sat beside the door that was getting painted.	REL	AUX <sub>PASS</sub> -PART	INTR
S11	e	/ε/ <i>beta</i> /βeta/ 'bitter'	The cherries in our garden have never been bitter.		(free-form)	
S12	ea	/ea/ <i>Beakja</i> /beakja/ 'books'	I'm happy that we can read books together.	COMP	MOD-INF	DO-IO
S13	ea	/əɪa/ <i>vea</i> /fəɪa/ 'four'	Mom wanted to cook four big hams for supper today.		(free-form)	
S14	eaj	/εaɪ/ <i>Kjeaj</i> /kjeaɪ/ 'cows'	If he could have sold his cows, he would be a very rich man.	COND	AUX-MOD-INF	DO-IO
S15	ee	/əɪ/ <i>Beete(n)</i> /bɛɪtə(n)/ 'beets'	My parents wanted to plant beets between the corn.		(free-form)	
S16	ei	/εɪ/ <i>Weit(e)</i> /veit(ə)/ 'wheat'	The field was green because the farmer had seeded wheat.	CAUS	AUX <sub>PERF</sub> -PART	DO-IO
S17	f	/f/ <i>Foahra</i> /fɔara/ 'driver'	Could you say to the driver that he should drive slowly?		(free-form)	

<i>ID</i>	<i>Sound</i>	<i>Word</i>	<i>Sentence</i>	<i>Clause</i>	<i>Construction</i>	<i>Transitivity</i>
S18	g	/g/	<i>Goade(n)</i> /gɔadə(n)/ 'garden'	I'm always the one that has to weed the garden!	REL	MOD-INF DO-IO
S19	g	/ɣ/	<i>Foagel</i> /fɔagəl/ 'bird'	I'm not sure if this grey bird can understand English.		(free-form)
S20	gj	/gʲ/	<i>Migje(n)</i> /mɪgʲə(n)/ 'mosquitos'	The mosquitos didn't know that they were supposed to stay outside.	COMP	AUX-MOD-INF INTR
S21	h	/h/	<i>Hoat</i> /hɔat/ 'heart'	He has a soft heart.		(free-form)
S22	i	/ɪ/	<i>witt</i> /vit/ 'white'	If we had made white bread, then we could have had <i>faspā</i> .	COND	AUX <sub>PERF</sub> -PART DO-IO
S23	ia	/ia/	<i>hia</i> /hia/ 'here'	My uncles and aunts are all here.		(free-form)
S24	ie	/i/	<i>wiet</i> /vit/ 'far'	The girls stayed far away because there was smoking (going on).	CAUS	AUX <sub>PASS</sub> -PART INTR
S25	j	/j/	<i>Joah</i> /jɔa/ 'year'	Every year has 365 days.		(free-form)
S26	jch	/ç/	<i>Laicha</i> /laça/ 'holes'	The road had big holes that could have been dangerous.	REL	AUX-MOD-INF INTR
S27	k	/k/	<i>Koa</i> /kɔa/ 'car'	If their old car gets sold, then they'll come visit us.	COND	AUX <sub>PASS</sub> -PART INTR
S28	kj	/kʲ/	<i>Kjoasche(n)</i> /kʲɔʃə(n)/ 'cherries'	They were sorry that they had eaten so many cherries.	COMP	AUX <sub>PERF</sub> -PART DO-IO
S29	l	/l/	<i>lot</i> /lot/ 'late'	We have often visited with our neighbours until late at night.		(free-form)
S30	lj	/li/	<i>Eelj</i> /əlɪ/ 'oil'	We don't have any oil that we can use in a lamp.	REL	MOD-INF INTR
S31	m	/m/	<i>Mon(d)</i> /mon(t)/ 'moon'	Can you see the moon early in the evening?		(free-form)
S32	n	/n/	<i>Noba</i> /noba/ 'neighbour'	Our neighbours want to earn money because they're going to build a new house.	CAUS	MOD-INF DO-IO
S33	ng	/ŋ/	<i>Hunga</i> /hɔŋa/ 'hunger'	Fresh bread strikes hunger dead.		(free-form)
S34	nj	/ɲ/	<i>Kjinja</i> /kɪɲa/ 'children'	We played with all of the children that had come.	REL	AUX <sub>PERF</sub> -PART INTR

<i>ID</i>	<i>Sound</i>	<i>Word</i>	<i>Sentence</i>	<i>Clause</i>	<i>Construction</i>	<i>Transitivity</i>
S35	o /ɔ/	<i>Botta</i> /bɔta/ 'butter'	They ate their bread without much butter.		(free-form)	
S36	o /o/	<i>Wota</i> /vota/ 'water'	The farmer's wife wanted to be sure that the water got boiled first.	COMP	AUX <sub>PASS</sub> -PART	DO-IO
S37	oa /ɔa/	<i>Boat</i> /bɔat/ 'beard'	The old man cut off his beard because it had turned grey.	CAUS	AUX <sub>PERF</sub> -PART	INTR
S38	oaj /ɔaj/	<i>Boaj</i> /bɔaj/ 'mountains'	If we want to see snow, then we should drive to the mountains.	COND	MOD-INF	DO-IO
S39	öa /œa/	<i>Woage(n)</i> /vɔage(n)/ 'wagon'	My brother climbed off of the wagon and went into the house.		(free-form)	
S40	oo /əʊ/	<i>Foot</i> /fəʊt/ 'foot'	George's foot was sore because he hadn't wanted to buy new shoes.	CAUS	AUX-MOD-INF	DO-IO
S41	p /p/	<i>Poa</i> /pɔa/ 'pair'	We have a pair of grey horses and two old wagons.		(free-form)	
S42	r /r/	<i>Rot</i> /rot/ 'advice'	That was the only advice that we were given.	REL	AUX <sub>PASS</sub> -PART	DO-IO
S43	s /z/	<i>Sot</i> /zot/ 'seed'	My parents took the seed out and gave it to me.		(free-form)	
S44	ss /s/	<i>Massa</i> /masa/ 'knife'	If the boys will help, maybe dad will let them use the knife.	COND	MOD-INF	INTR
S45	sch /ʃ/	<i>schoap</i> /ʃɔap/ 'sharp'	If the knife has gotten dull, the boys can help make it sharp again.	COND	AUX <sub>PERF</sub> -PART	INTR
S46	t /t/	<i>tohm</i> /tom/ 'tame'	Do you know if the blue birds are tame?		(free-form)	
S47	ts /ʃs/	<i>Zocka</i> /ʃɔka/ 'sugar'	If we could have baked today, then you would have had sugar cookies.	COND	AUX-MOD-INF	INTR
S48	tsch /ʃʃ/	<i>Dietsch</i> /difʃ/ 'German'	Our relatives could visit with us because German is still spoken here.	CAUS	AUX <sub>PASS</sub> -PART	DO-IO
S49	u /ʊ/	<i>Buck</i> /bʊk/ 'stomach'	Dad says that we shouldn't have rubbed her stomach.	COMP	AUX-MOD-INF	DO-IO
S50	ü /u/	<i>Huus</i> /hus/ 'house'	Dad wanted to see that the house got built properly.	COMP	AUX <sub>PASS</sub> -PART	INTR
S51	ua /oa/	<i>Wuat</i> /voat/ 'word'	At the beginning the kids didn't want to say a word.		(free-form)	



<i>ID</i>	<i>Sound</i>	<i>Word</i>	<i>Sentence</i>	<i>Clause</i>	<i>Construction</i>	<i>Transitivity</i>
S52	üa /øa/	<i>Buak</i> /boak/ 'book'	The book that my mom had wanted to give her was very beautiful.	REL	AUX-MOD-INF	DO-IO
S53	uj /ɔɪ/	<i>fuj</i> /fɔɪ/ 'phooey!'	"Phooey!", said Abram, and spat out the watermelon seeds.		(free-form)	
S54	w /v/	<i>woat</i> /vɔat/ 'will (be)'	If a new store gets built, then the old one will be sold.	COND	AUX <sub>PASS</sub> -PART	DO-IO
S55	zh /ʒ/	<i>Bockelzhann</i> /bøkəl'ʒan/ 'tomato'	We helped grandma pick tomatoes because she can't see very well.	CAUS	MOD-INF	INTR

Table 16. Design of the *Fibel*.

Although these revisions and review contributed significantly to the development of the *Fibel*, they did not eliminate all shortcomings of the final design. While a critical assessment of the *Fibel* is undertaken in Section 4.2.3.5, the following sections first turn their attention to the development of the written *Fibel* into a collection of spoken language documentation and the subsequent creation of a linguistically annotated, time-aligned corpus on the basis of these materials.

#### 4.2.3 *Developing the Fibel Corpus*

Developing the *Fibel* from its initial form as a series of written, English-language word and sentences into a collection of spoken, Plautdietsch-language equivalents involved several separate tasks. This began with recording contributors' oral translations of these prompts into their respective varieties of Saskatchewan Plautdietsch (§4.2.3.1). To be more amenable to later reuse, these recordings were then transcribed and translated (§4.2.3.3), and coding subsequently added to associate sections of these materials with corresponding prompt words, sentences, and embedded linguistic variables (§4.2.3.4). The following sections describe each of these stages in developing the written *Fibel* into a spoken corpus before summarizing the major features of this approach (§4.2.3.5).

##### 4.2.3.1 *Contributions and contributors*

The approach to developing the *Fibel* into a reusable language resource outlined above relies on individuals contributing translations of English prompt words and sentences into their respective varieties of Plautdietsch. This raises important questions about these contributors—how they were identified (and how this relates both to methodological recommendations on the sampling of speaker populations and to the realities of demographic skew found in the wake of significant language shift; cf. §1.2.1, §1.2.4, §1.2.5, §2.6), what information each contributor provided, and how concerns over acknowledgment, anonymity, and appropriate reuse of information were addressed. These issues are therefore taken up in greater detail in this section.

As noted in Chapter 1, there is some diversity in the perspectives on sampling commonly encountered in quantitative sociolinguistics (§1.2.1), corpus linguistics (§1.2.5), and documentary linguistics (§1.2.4). In the former two fields, in particular, considerable emphasis

has been placed on the importance of satisfying *a priori* sociodemographic categories of interest, such that the resulting samples of the speaker population are evenly balanced with respect to these features. The preceding discussion has noted both the value of such approaches in ensuring systematic representation of what are presumably significant sections of the speaker population; as well as their potential drawbacks in the present context, where the effects of language shift and the intimate nature of the remaining domains of vernacular language use present particular challenges for common methods of sampling advocated in this literature. These approaches contrast with others that are more prominent in documentary linguistics, where more opportunistic forms of sampling are sometimes advanced as being more suitable in situations of advanced language endangerment, where even representation of speakers across a predetermined set of sociodemographic categories may be difficult to achieve (cf. Woodbury 2003).

Given the sharp contrast between these two approaches to sampling, this study took a mediating approach, adopting a form of snowball sampling similar to what Milroy (1987) proposes for variationist sociolinguistic research into the structure of social networks. In this model, each contributor was asked if he or she might know of anyone else in the community who may also be willing to contribute a translation of the *Fibel*. Such contacts facilitated by members of the local community constitute ‘friend-of-a-friend’ introductions similar to the kind advocated by Milroy, and had the advantage of ensuring a degree of familiarity and mutually understood social connection between potential contributors and a possibly otherwise unknown interlocutor. While such sampling is still ‘opportunistic’ in the sense of Woodbury (2003) and is not guaranteed to produce even balance across a set of given demographic categories or to avoid undue bias towards certain social networks, it was nevertheless able to extend invitations to a broad segment of the speaker population on the basis of existing relationships of trust, as is critical in the present sociolinguistic context. Such potential contributors were identified here in several ways: through personal contacts in Saskatchewan Valley communities, through the advice of local advisors and introductions facilitated by them, and through participation in community language events (e.g., social evenings held regularly in Saskatoon, where individuals from Saskatchewan Valley communities gathered with the purpose of speaking Plautdietsch with one another). Especially through the latter meetings, where several dozen individuals expressed an interest in assisting with the *Fibel*, it was possible to ensure a broad base of representation

from communities throughout the Saskatchewan Valley and avoid potential bias in sampling that might otherwise have unduly favoured individuals from the social networks of the author and the local advisory group. Further introductions facilitated by these first contributors broadened the representation of Mennonite communities in the Saskatchewan Valley, helping to ensure more adequate representation of even smaller settlements in the region.

Such assistance also led to contact with members of another, non-Mennonite Plautdietsch speech community in central Saskatchewan.<sup>41</sup> A small, Catholic Plautdietsch-speaking community exists in the area of Humboldt, Saskatchewan, approximately 100 kilometres east of the Saskatchewan Valley. The members of this community are the descendants of settlers in St. Peter's Colony, a Catholic settlement established in central Saskatchewan in 1903 (Hepp 1998, Paproski 2011). These colonists emigrated to Canada from the Mariupol region of Ukraine, with their forebears having left northern Poland in the early 19th century (McIver 1996, Epp 1996: 2). This Plautdietsch-speaking community persisted as a distinct linguistic minority among the generally High German-speaking population of St. Peter's Colony, settling primarily in the southwestern corner of the colony south of Carmel, Saskatchewan, and near Burr, Saskatchewan (Hepp 1998: 9–10, Paproski 2011: 87–88).<sup>42</sup> Despite the mutual intelligibility of their respective varieties of Plautdietsch and the relatively small geographical distance separating them, there appears to have been little historical contact between these Mennonite and Catholic speaker communities, either in Ukraine or in Canada. Outside of the few words and traditional sayings recorded in an appendix to McIver (1996), the Catholic Plautdietsch varieties spoken in Saskatchewan appear to be entirely undocumented. The participation of several Catholic Plautdietsch speakers in the development of the *Fibel* thus presented an opportunity not only to contribute to the language resources available to members of this community, but also to consider the relationship of these varieties to Mennonite forms of Plautdietsch spoken in Saskatchewan.

---

41 Particular thanks are due to Leonard Doell, Henry Kloppenburg, Q.C., and the late Reuben Epp for their help in facilitating introductions with members of the Catholic Plautdietsch-speaking community in Saskatchewan.

42 There are reportedly also speakers of related Catholic Plautdietsch varieties in the area of St. Joseph's Colony (near present-day Leipzig and Balgonie, Saskatchewan, approximately 150 kilometres west of the Saskatchewan Valley; cf. Hepp 1998: 10), although it has not been possible to make contact with any members of this community to date.

Both the Mennonite and Catholic contributors of Plautdietsch translations of the *Fibel* are summarized in Table 17 below. Along with the unique identifier and year of birth of each of the 49 contributors, several additional pieces of information (e.g., both the contributor's and his or her parents' denominational affiliation and place of birth) are also given in the table; these details are discussed in more detail in the following section.

<i>ID</i>	<i>DOB</i>	<i>Denom.</i>	<i>POB</i>	<i>Parents' POB</i>
F00	1935 (78)	MC (MC)	<i>Rabbit Lake, SK</i>	UKR (Ch.)
F01	1922 (91)	RC (RC)	<i>Carmel, SK</i>	UKR (Ma.)
F02	1938 (75)	B (B)	Aberdeen	CAN (WR)
F03	1926 (87)	MC (MC)	Langham	USA (NE, MN)
F04*	1925 (88)	RC (RC)	<i>Carmel, SK</i>	CAN (SK), UKR (Ma.)
F05	1933 (80)	OC (OC)	Kronsthal	CAN (MB)
F06	1928 (85)	MC (MC)	Hepburn	UKR (< Ch.)
F07	1933 (80)	MC (MC)	Osler	UKR (Ch.)
F08	1934 (79)	MC (MC)	Osler	UKR (Ch.)
F09	1939 (74)	MC (MC)	Osler	CAN (SK), UKR (< Ch.)
F10*	1940 (73)	MC (MC)	Hague	CAN (SK)
F11	1940 (73)	MC (OC)	Hague	CAN (SK)
F12	1920 (93)	RC (RC)	<i>Carmel, SK</i>	UKR (Ma.)
F13*	1934 (79)	MC (MC)	<i>Fiske, SK</i>	UKR (Ch.)
F14*	1958 (55)	MC (B)	<i>Saskatoon, SK</i>	CAN (SK)
F15	1941 (72)	E (MC)	Osler	CAN (SK), UKR (< Ch.)
F16*	1934 (79)	RC (RC)	<i>Humboldt, SK</i>	CAN (SK), UKR (Ma.)
F17	1944 (69)	MC (B)	Osler	CAN (SK)
F18	1922 (91)	MC (OC)	Hague	CAN (WR)
F19	1934 (79)	FEBC (MC)	<i>Gretna, MB</i>	CAN (WR)
F20	1930 (83)	MC (MBr)	Rosthern	UKR (Mo.)
F21	1939 (74)	CMC (OC)	Hochfeld	CAN (MB)
F22	1928 (85)	B (B)	Reinfeld	CAN (SK)
F23	1936 (77)	MC (MC)	Rosthern	UKR (< Mo.)
F24	1936 (77)	MC (MC)	<i>Rabbit Lake, SK</i>	UKR
F25	1937 (76)	MC (MC)	<i>Elbow, SK</i>	UKR
F26	1934 (79)	RC (RC)	<i>Carmel, SK</i>	UKR (Ma.)
F27	1927 (86)	MC (B)	Hague	CAN (WR)
F28	1932 (81)	MC (OC)	Warman	CAN (SK)
F29	1946 (67)	MC (OC)	Warman	CAN (SK, WR)
M00	1929 (84)	MC (MC)	<i>Didsbury, AB</i>	CAN (AB)
M01	1925 (88)	MC (MC)	Schlorrendarp	UKR (Ch.)
M02	1937 (76)	MC (MC)	Rosthern	CAN (SK), UKR (Ch.)
M03	1944 (69)	MC (OC)	Neuhorst	CAN (SK)

<i>ID</i>	<i>DOB</i>	<i>Denom.</i>	<i>POB</i>	<i>Parents' POB</i>
M04	1942 (71)	MC (MC)	Osler	UKR (< Ch.)
M05	1946 (67)	MC (MC)	Warman	UKR (< Ch.)
M06	1955 (58)	MC (OC)	Neuhorst	CAN (SK)
M07	1937 (76)	MC (OC)	Hochfeld	CAN (MB)
M08	1922 (91)	MC (OC)	Neuanlage	CAN (SK)
M09*	1937 (76)	CMC (MC)	Langham	USA (NE), UKR
M10	1931 (82)	MC (MC)	Dalmeny	UKR (< Ch.)
M11	1921 (92)	MC (OC)	Hepburn	UKR (Ch., 1870)
M12	1934 (79)	E (MC)	<i>Saskatoon, SK</i>	CAN (MB)
M13*	1928 (85)	FEBC (MC)	<i>Lowe Farm, MB</i>	CAN (WR)
M14	1937 (76)	B (B)	Aberdeen	CAN (WR, SK)
M15	1917 (96)	MC (OC)	Hague	CAN (MB, SK)
M16	1926 (87)	MC (OC)	Blumenheim	CAN (WR)
M17	1940 (73)	MC (MC)	Reinland	CAN (SK)
M18	1948 (65)	MC (OC)	Neuhorst	CAN (SK)

*Table 17. Contributors to the Fibel Corpus.* Contributors' gender is indicated in the first letter of each anonymous identifier, and contributor ages are given as of the time of their participation. Starred contributors are only sparsely represented in the corpus. Denominational affiliations for contributors and their parents (in parentheses) use the following conventions: B = Bergthaler, CMC = Chortitzer Mennonite Conference, E = Evangelical, FEBC = Fellowship of Evangelical Bible Churches, MBr = Mennonite Brethren, MC = Mennonite Church Canada, OC = Old Colony, RC = Roman Catholic. Contributors' places of birth outside of the Saskatchewan Valley are presented in italics. Contributors' parents' places of birth are divided by country and region: CAN = Canada, with MB = Manitoba (East or West Reserve), SK = Saskatchewan, and WR = West Reserve, Manitoba; UKR = Ukraine, with Ch. = Chortitza Colony, <Ch. = daughter colony of the Chortitza Colony, Mo. = Molochnaya Colony, Ma. = Mariupol region; and USA = United States, with MN = Minnesota and NE = Nebraska.

This summary underscores the considerable diversity in emigrational, denominational, and settlement background noted among contributors, who represent the descendants of participants in every significant wave of Mennonite migration into Saskatchewan.<sup>43</sup> Part of this internal diversity has developed more recently; there is evidence of widespread intergenerational shift in denominations among the Mennonite contributors, particularly from the Old Colony and Bergthaler *Gemeinden* to Mennonite Church Canada (and, to a lesser extent, from Mennonite

<sup>43</sup> Several contributors have historical ties to more than one wave of emigration, as when one parent was born in Canada and the other in Ukraine (e.g., F09, F15, M02). While most contributors are the children of either recently arrived *Russländer* or of the first or second generations of Canadian-born *Kanadier* Mennonites, one contributor (M11) is exceptional in being the child of participants in the original 1870s migration to Canada.

Church Canada to the Chortitzer Mennonite Conference and Fellowship of Evangelical Bible Churches). These contributors' places of birth also encompass a broad sample of the Mennonite communities in the Saskatchewan Valley, albeit not a complete one. Several smaller villages in the region are not represented, in large part due to their historical depopulation through mass emigration to Latin America.<sup>44</sup> Among these contributors are also individuals who, while often long-term residents of the Saskatchewan Valley and part of the local speech community, were born outside of the valley proper (e.g., in Saskatoon, particularly in later years when hospital-assisted births became more common). In this respect, place of birth is perhaps a less adequate indicator of regional affiliation than where an individual was raised, but still arguably presents serviceable information for analysis (cf. §5.2).

There is also some imbalance in the proportional representation of genders and ages among these contributors. There are slightly more women (30, 25 of whom contributed full translations of the *Fibel*) in the group than men (19, 17 of whom contributed full translations), although this difference is relatively minor. More pronounced is the skewed distribution of ages: the youngest contributor, F14, was 55 years old at the time of participation, while the oldest, M15, was 96. The mean and median age of contributors was 78.7 and 79, respectively, with a standard deviation of 8.7 years. Figure 4 summarizes the distribution of contributor ages. This age range is not atypical for the Saskatchewan Valley, where fluent speakers in their mid-fifties are increasingly uncommon.

---

44 Nevertheless, other, somewhat larger communities such as Waldheim, Langham, and Laird are also not well represented, and would benefit from further attention in continued research.

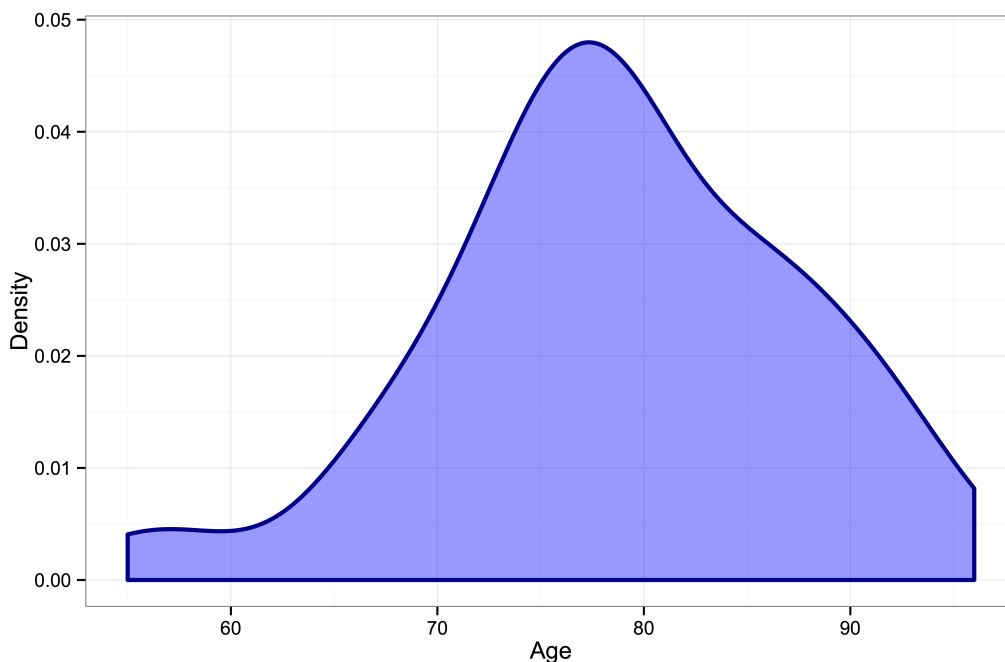


Figure 4. Density plot of contributor ages.

A final note is required on the use of unique identifiers throughout this study to refer to individual contributors. Although all contributors to the *Fibel* indicated that they would be willing to be acknowledged by name, it was assured that any additional, personal information that they provided about themselves or their family history would only be used publicly in anonymized form. As elsewhere in linguistics (cf. Childs, van Herk & Thorburn 2011), there is a degree of tension here between the requirements of adequate acknowledgement and assured anonymity. Any public release of personal details in non-anonymized form, for instance, may reveal enough information to allow the same contributors to be identified in materials for which anonymity was guaranteed. By the same token, a wholesale lack of recognition for individual contributors' part in assisting in the development of these language resources seems inappropriate, particularly for the uses in local communities for which these materials are intended. In this study, a compromise was reached by ensuring that contributors are identified only by randomly assigned anonymous identifiers here, and that all demographic information is presented at a level of detail that would make it difficult to identify individual contributors. To acknowledge their involvement, contributors' names are included in the preface to this study among the much larger group of individuals from Saskatchewan Plautdietsch-speaking



communities who have contributed to this work. Although this preserves contributors' anonymity while acknowledging their involvement, it is still necessary to ensure that later reuse of these materials is accompanied by the condition that speaker anonymity be preserved similarly, taking steps accordingly to avoid the unsanctioned release of any sensitive information.<sup>45</sup>

#### 4.2.3.2 *Tasks and recording*

Each of the contributors acknowledged in the preceding section assisted in translating the *Fibel* into their variety of Plautdietsch. In general, the meetings in which these translations took place in the homes of the contributors or in mutually agreed-upon public places and involved three steps:

1. After introductions and general visiting, the aims and methodology of the project were discussed, such that the contributors could decide in an informed manner whether or not to participate. An information sheet (in English) describing the project and providing the researcher's contact information was given to the contributors, so that any questions or concerns about the project or their involvement could be addressed in the future;
2. After setting up the recording equipment (see below), the contributor reviewed a printed copy of the *Fibel* together with the researcher, proceeding page by page through the document. Typically, the researcher would read the prompt word and sentence in English, and contributors were asked to translate these into Plautdietsch as they learned to speak it at home, repeating each of their translations twice;
3. Once the *Fibel* had been reviewed, a personal information form was provided to the contributor to complete and return at his or her convenience (or during the meeting with the assistance of the researcher, if requested). Contributors were assured that they would receive copies of their recordings as soon as these could be processed (cf. §4.2.3.3).

In a few cases, contributors indicated that they would prefer not to complete their translation of the *Fibel*. These individuals (F10, M09) are among those for whom only sparse representation is

---

<sup>45</sup> To ensure that the results of this work remain accessible to local communities and respect the aforementioned conditions on access and reuse, arrangements have been made for copies of all materials from this study to be deposited with the archives of the Mennonite Historical Society of Saskatchewan, a community-based institution with the mandate of providing long-term preservation and appropriate access to such information.

available in the final *Fibel* Corpus. The remaining individuals whose contributions to the *Fibel* Corpus are similarly circumscribed (i.e., F04, F13, F14, F16, and M13) were present for a translation of the *Fibel* by another speaker, but declined to provide a full translation of their own. In all such cases, these individuals' decision to limit their participation was respected, and no pressure was placed on them to be involved beyond the level they felt was appropriate.

As part of their involvement in the development of the *Fibel*, all contributors were also asked if they would be willing to provide additional information about themselves and their family history, in order to better understand the different varieties of Plautdietsch in Saskatchewan and their relationship to the history of local communities. This information, summarized in Table 17 above, included each contributor's name, gender, occupation, educational background (i.e., whether completing formal education before grade eight, after grade eight, at high school, with professional training, or with university), date and place of birth, current denominational affiliation, and any places he or she lived for three or more years. Further information was requested on the contributor's family history, including the names, dates and places of birth, occupation, and denominational affiliation of his or her parents. Also included was a two-page survey on the languages that each contributor knew; a self-assessment of how often he or she spoke each language and how well he or she could understand, speak, and read each language; and the individuals with whom and contexts in which the contributor spoke each language. All of this information was gathered on a form (reproduced in Appendix D) that explained the purpose of requesting these details and provided assurance that this information would only ever be presented publicly in anonymized form. Contributors generally had little difficulty with the personal information sheets, but reported finding the language use forms more challenging to complete, given the amount of information requested and degree of reflection required to respond accurately. The results of this survey varied substantially in both quality and completeness between contributors, and are therefore drawn on only sparingly in this study.

Following the recommendations outlined in Section 1.2.4 on recording spoken language, care was taken to choose recording equipment and associated standards that supported the broader goal of this project to produce a lasting record of local varieties of Plautdietsch that is amenable to future reuse. All recordings of translations of the *Fibel* were produced in digital form using non-proprietary, uncompressed audio formats at sample rates and with sample sizes

recommended for representing speech (here, uncompressed linear PCM audio in .wav format, recorded at 48 KHz with 24-bit samples). After several rounds of preliminary testing (both for technical quality and comfort on the part of contributing speakers), it was decided that all *Fibel* recordings would be made with Countryman E6i omnidirectional earset microphones (professional-grade microphones that offer a flat frequency response pattern and low self-noise levels; cf. Lee 2013) and a solid state audio recorder (either an Edirol R-09HR with a Shure MixPre pre-amp, or, later, a Sound Devices 702). To reduce the amount of recording equipment needed and to limit its impact on contributors' sense of ease in this context, no video recordings were made of these meetings. The use of an earset microphone in particular proved well suited to this context, and several contributors later commented that the microphone was not only comfortable to wear and easily forgotten about in the course of a meeting, but also much less obtrusive and intimidating than a stand-mounted or hand-held microphone may have been.<sup>46</sup>

When proceeding through the *Fibel* with a contributor, the researcher attempted to use Plautdietsch sparingly (outside of short back-channel responses such as *jo* 'yes' intended to encourage contributors to continue speaking in the language, even with a younger interlocutor), favouring English where possible to attempt to avoid undue influence from the researcher's own L2 variety of Plautdietsch on the varieties spoken by contributors. If the prompt item had accidentally been skipped over, the researcher would often attempt to remind the contributor and ask for a translation (e.g., "How would you say (prompt item) in Plautdietsch again?"). Similarly, if a second repetition was not provided for a particular item, the researcher would encourage contributors to repeat their previous response again (e.g., "How would you say that again?", "How did you say that?").<sup>47</sup> On the whole, few contributors deviated from this procedure, or expressed any concern over the difficulty of the task (outside of repeated comments on the necessity of adjusting the order of words between English prompt sentences and Plautdietsch translations, or occasional requests to leaf through the *Fibel* briefly while

---

46 Only in session 2011-08-07 (01) did the contributors (\*F14 and M06) request that the earset microphone not be worn, but rather held in front of the individual providing translations. This request was not difficult to accommodate and resulted in only slightly diminished audio quality.

47 In a few rare cases, an entire page of the *Fibel* was accidentally omitted while completing a translation. While unfortunate, redundant occurrences of the same variables in other sections of the *Fibel* often mitigated the negative effects of such unintentional gaps.

visiting before the translation session).<sup>48</sup> The recordings that resulted from these meetings are listed in Table 18, which notes the name and duration of each recording (with multiple parts of the same recording session indicated in parentheses, where applicable), as well as the recording device used, the location of the recording, and the contributors involved.

<i>Recording</i>	<i>Device</i>	<i>Location</i>	<i>Contributor(s)</i>	<i>Length</i>
2011-07-22 (01)	Edirol	Middle Lake, SK	F01	53m58
2011-07-22 (02)	Edirol	Middle Lake, SK	F01	02m46
2011-07-23	Edirol	Warman, SK	F21 (M09*)	38m20
2011-07-23 (01)	Edirol	Saskatoon, SK	F10* (M07)	10m39
2011-07-23 (02)	Edirol	Saskatoon, SK	M07 (F10*)	22m43
2011-07-23 (03)	Edirol	Saskatoon, SK	M07 (F10*)	26m23
2011-07-23 (01)	Edirol	Saskatoon, SK	M16	38m29
2011-07-23 (02)	Edirol	Saskatoon, SK	M16	06m29
2011-08-02	SD702	Osler, SK	M18	52m55
2011-08-02 (01)	SD702	Saskatoon, SK	M04	03m32
2011-08-02 (02)	SD702	Saskatoon, SK	M04	33m45
2011-08-03 (01)	SD702	Warman, SK	M14 (F02)	19m11
2011-08-03 (02)	SD702	Warman, SK	F02 (M14)	26m25
2011-08-03 (01)	SD702	Osler, SK	F09 (M02)	27m43
2011-08-03 (02)	SD702	Osler, SK	M02 (F09)	24m13
2011-08-03 (01)	SD702	Saskatoon, SK	M10 (F13*)	69m56
2011-08-03 (02)	SD702	Saskatoon, SK	M10 (F13*)	04m40
2011-08-03 (01)	SD702	Saskatoon, SK	F06 (M01)	02m34
2011-08-03 (02)	SD702	Saskatoon, SK	F06 (M01)	14m03
2011-08-04 (01)	SD702	Saskatoon, SK	M15	24m42
2011-08-04 (02)	SD702	Saskatoon, SK	M15	11m46
2011-08-04	SD702	Saskatoon, SK	F18	40m39
2011-08-04 (01)	SD702	Osler, SK	F29 (M05)	43m04
2011-08-04 (02)	SD702	Osler, SK	M05 (F29)	25m06
2011-08-05 (01)	SD702	Saskatoon, SK	F19 (M13*)	26m15
2011-08-05 (02)	SD702	Saskatoon, SK	F19 (M13*)	19m07
2011-08-05	SD702	Saskatoon, SK	M11	34m58
2011-08-05	SD702	Saskatoon, SK	F25	36m50
2011-08-07	SD702	Neuhorst, SK	M06 (F14)	48m08
2011-08-08 (01)	SD702	Saskatoon, SK	F07	21m16
2011-08-08 (02)	SD702	Saskatoon, SK	F07	18m55
2011-08-08	SD702	Saskatoon, SK	F12	48m47
2011-08-08	SD702	Osler, SK	F08	30m41
2011-08-09 (01)	SD702	Saskatoon, SK	F20 (M00)	54m52
2011-08-09 (02)	SD702	Saskatoon, SK	M00 (F20)	48m16
2011-09-13 (01)	SD702	Saskatoon, SK	F27 (M08)	35m17

48 One exception is session 2011-10-28 with F23, who asked that the *Fibel* first be reviewed in full, then translated.

<i>Recording</i>	<i>Device</i>	<i>Location</i>	<i>Contributor(s)</i>	<i>Length</i>
2011-09-13 (02)	SD702	Saskatoon, SK	M08 (F27)	30m16
2011-09-13 (01)	SD702	Saskatoon, SK	F03	30m06
2011-09-13 (02)	SD702	Saskatoon, SK	F03	09m33
2011-09-13 (03)	SD702	Saskatoon, SK	F03	10m33
2011-09-15 (01)	SD702	Saskatoon, SK	F00, F24	36m49
2011-09-15 (02)	SD702	Saskatoon, SK	F00, F24	03m50
2011-09-16	SD702	Humboldt, SK	F26 (F04*, F16*)	56m28
2011-10-26	SD702	Warman, SK	F15, M12	46m14
2011-10-27 (01)	SD702	Osler, SK	F17, M03	28m57
2011-10-27 (02)	SD702	Osler, SK	F17, M03	17m56
2011-10-27	SD702	Warman, SK	F28	41m47
2011-10-28 (01)	SD702	Osler, SK	F23	04m09
2011-10-28 (02)	SD702	Osler, SK	F23	36m32
2011-10-28 (03)	SD702	Osler, SK	F23	14m13
2011-10-29	SD702	Osler, SK	F11, M17	57m29
2011-10-31	SD702	Warman, SK	F05	46m39
2012-10-18	SD702	Saskatoon, SK	F06, M01	59m52
2012-10-18	SD702	Warman, SK	F22	46m15
			<i>Total:</i>	27h05m01

*Table 18.* Contents of the *Fibel* Corpus. Contributors here include all individuals present for the *Fibel* task, with individuals in parentheses not providing a full translation in the given recording, and starred contributors being sparsely represented in the corpus overall.

Some variation is noted in the length of time that contributors took to complete the *Fibel* translation task. Several speakers (e.g., M14) were able to translate the entire *Fibel* in less than twenty minutes, while others (e.g., M10) took well over an hour. In most cases, these differences related to the amount of commentary that contributors provided to accompany their translations, rather than to any difficulties they experienced with the task itself.<sup>49</sup> Other differences between sessions may be of somewhat greater consequence, however, as when meetings were conducted with more than one Plautdietsch speaker present. Given the settings in which these meetings took place, it was not generally socially appropriate to ask one contributor to leave while another provided a translation. In some such cases, contributors opted to translate the *Fibel* one after another (*sequential translation*, as in session 2011-08-03 with F09 and M02), or to take turns

<sup>49</sup> With two of the recordings, however, technical problems led to parts of the recording being truncated: the last few minutes of session 2011-07-22 were lost due to a loss of power to the recording device, while a more significant portion of session 2011-08-03 (01) with F06 and M01 was also truncated, but later recorded again in session 2012-10-18.

translating individual pages of the *Fibel* (*simultaneous translation*, as in session 2012-10-18 with F06 and M01). This inevitably introduces an element of methodological variation into the study, raising questions as to the independence of individuals' responses.

Given that sessions representing all three translation conditions (i.e., independent, sequential, and simultaneous translations) are available, this issue can be treated as an empirical question: to what extent do individuals who provided sequential and simultaneous translations demonstrate convergence in their responses, and are any such differences from independent translations statistically apparent in the *Fibel* data? From a computational linguistic perspective, contributors' translations of sentences in the *Fibel* can be treated as sparse feature matrices containing counts of the occurrence of individual words. Applying a cosine distance measure to pairs of these matrices provides a sense of their similarity. If a pair of sentences have most of their contents in common, then their cosine distance will have a value closer to one, while sentences sharing fewer features will have cosine distance values closer to zero. Using these measures, it is possible to assess the mean similarity of all sentence translations provided by contributors in simultaneous and sequential sessions, and to compare this with the mean similarity of these contributors' responses and those of other, similar speakers (cf. §5.2). A Wilcoxon test reveals a statistically significant difference in these two means ( $W = 109$ ,  $Z = -2.9866$ ,  $p = 0.01319$ ), with joint translations being generally more similar than those provided by individual contributors on their own. These results could be interpreted in several ways: it could be that joint contributors' responses are more similar to one another because of the nature of the translation task itself, which may have encouraged convergent responses; or that these pairs of speakers show greater similarity because of other shared personal connections, and that this apparent 'convergence' is, in fact, representative of their regular individual speech patterns. However these results are best understood, they nevertheless provide serviceable information for the design of further studies, where such factors can perhaps be better controlled for in the arrangement of individual sessions, and, perhaps more importantly for the present study, allows for such skew to be taken into account when interpreting the results of later analysis.

Regardless of the recording context, the audio materials that resulted from each meeting were organized into sessions with consistent, platform-independent file names, as per current documentary linguistic recommendations for linguistic data management (Thieberger & Berez

2012). Where recordings were made in stereo, but only employed a single channel, the unused channel was removed and the recording saved in otherwise unedited form as monophonic audio with the same sample rate and size as the original. Both MD5 and SHA-1 checksums were subsequently generated for each recording, allowing any changes to the contents of the files to be identified computationally and addressed (e.g., in identifying and correcting data corruption encountered in any copies of these materials). Access copies of these preservation materials were then produced for use in later stages of annotation and copies of each session's recordings were made on audio CD to be returned to each of the contributors. Personal information sheets received from contributors were entered into a spreadsheet and saved as Unicode text in comma-separated value (CSV) format. All of the resulting materials—master recordings, checksums, access copies, and contributor and session metadata—were finally backed up on multiple devices, with regular checks for file integrity being conducted to prepare them for archiving with a trusted local institution (see fn. 45 above).

#### 4.2.3.3 *Transcription*

The material result of the process described above was a consistently organized collection of high-quality digital audio recordings and metadata, representing over 27 hours of meetings with Plautdietsch speakers in Saskatchewan communities. While these materials may already be of use in unannotated form to individuals with a general familiarity with Plautdietsch, access to the contents of these recordings might be improved for both community and academic users through further stages of written transcription and translation. As transcription involves decisions that may affect the range of end uses to which the final materials can readily be put, this section considers such choices relating to orthographies, transcription conventions, and methods for associating written information with audiovisual recordings in greater detail.

With no single orthographic standard for Mennonite Plautdietsch being accepted across all diasporic communities (cf. Nieuweboer 1998), it was decided that the Plautdietsch adaptation of the standard Sass orthography for *Niederdeutsch* outlined in Epp (1996) would be used here. The decision to begin with a single spelling system was made with the aim of supporting consistent search and retrieval in transcribed materials. Were multiple orthographies applied or “dialect respelling” and “eye-dialect” transcription used to represent variation, it would be

essentially impossible to retrieve all instances of a particular item from the corpus without knowing the range of possible orthographic variants in advance or applying higher-level forms of annotation to each resource (cf. Preston 1985, Tagliamonte 2006). While this is not the only orthography in use in the Saskatchewan Valley (cf. Fehr 1993, 2001, 2006; Driedger 2011), it has the advantage of providing clearly defined spelling conventions and extensive accompanying lexical documentation that can be drawn on to ensure consistent transcription. Although no concerns were raised by members of the Saskatchewan Valley community about the use of this orthography here (perhaps in part because of the large number of conventions it shares with other, locally adopted systems), were such issues to arise, it would not be difficult to transliterate materials in the Sass-Epp orthography to another system automatically.

In addition to selecting consistent orthographic conventions for this project, it was also necessary to decide on transcription conventions that could represent the features of spoken language found in these recordings. Appendix E gives an overview of the conventions adopted for annotating the *Fibel* Corpus materials. Importantly, no representation of overlapping speech is found in these conventions. Instead, overlap is indicated implicitly through time alignment in the transcripts (i.e., by intersecting ranges of start and end times), rather than through any particular written transcription conventions. Including timestamps for each utterance not only reduces the overall number of transcription conventions required to capture features of spoken language (e.g., by eliminating conventions associated with overlapping speech, which can be accomplished here without the use of in-text symbols), but also provides immediate access to transcribed segments of the original recordings.

Time-aligned transcription was carried out using ELAN, an open-source documentary linguistic software tool developed to annotate audiovisual materials (Brugman & Russel 2004). In ELAN, transcripts are represented as sets of textual annotations having particular start and end times in an associated recording. These annotations are arranged hierarchically into tiers, which provide additional information about these annotations' type (e.g., transcribed text, translation, note, etc.) and contents (e.g., indicating the speaker associated with this annotation), and, optionally, place constraints on their organization (e.g., requiring that all free translations be associated with some portion of transcribed text, thus preventing 'orphaned' free translations from littering the transcript when corresponding sections of transcribed text are edited or



removed). ELAN facilitates the creation of time-aligned annotations within such constraints, thus ensuring that the resulting transcripts are consistently structured and amenable to further computational processing. With documentary linguistics rapidly converging on ELAN as the *de facto* standard for audiovisual annotation (cf. Berez 2007), and with ELAN storing its transcripts in a non-proprietary, Unicode-based XML format suitable for long-term preservation, its selection as the primary annotation tool for this corpus was in line with several of the goals of this project, and positioned this corpus to benefit from future developments in documentary linguistics as associated tools and workflows for ELAN-encoded corpora are developed.

For this corpus development, sets of tiers were defined in ELAN with a consistent range of types (e.g., ‘text’, ‘translation’, ‘note’, ‘coding’, etc.). Not only did this provide the consistency in annotation alignment noted above, but, importantly, also prepared annotations of these different types to be identified and extracted automatically at later points in corpus development. For each speaker in a recording, one such set of tiers was created, using the unique identifier for that speaker as part of the corresponding tier names. Thus, the ELAN transcripts for a session with contributor F05 contained a time-aligned tier named ‘F05’ (of the type ‘chunk’, containing transcriptions of all utterances that this contributor spoke in the recording), as well as subordinate tiers named ‘F05-Translation’ (of the type ‘translation’, containing a free translations in English of the associated utterance), ‘F05-Notes’ (of the type ‘notes’, containing general commentary on the associated utterance), ‘F05-Coding’ (of the type ‘coding’, containing formal coding of linguistic variables in the associated utterance; cf. §4.2.3.4 below), and ‘F05-Postprocess’ (of the type ‘subchunk’, containing time-aligned annotations representing sections of text that should be subjected to some form of postprocessing in the future before the corpus is released for public use—most often, omitting sections of the recording that contained personally identifiable information, such as personal names). Each recording in the corpus was annotated according to this scheme, ensuring that all Plautdietsch-language utterances in the corpus were fully transcribed.<sup>50</sup>

This phase of corpus development produced time-aligned annotations for all materials in the collection, represented in consistently organized, non-proprietary formats and adhering to the

---

50 Special thanks are due to Adrienne Findlay and Chelsea Cox, who assisted in segmenting corpus recordings into utterances for later transcription in ELAN.

transcription conventions outlined above. As expected, these layers of textual annotation, directly associated with sections of the corresponding recordings, presented a significant enrichment of the collected documentation. Whereas the original, predominantly Plautdietsch-dominant materials were only accessible to those able to understand spoken Plautdietsch, and only searchable at the level of the session (e.g., to retrieve all recordings in which a given speaker was present, but not any of their utterances), the newly annotated corpus was both now open to reuse by non-Plautdietsch-speaking audiences and to searches at the level of individual words and utterances for each contributing speaker. This enabled a considerably broader range of uses of the *Fibel* Corpus, including the analytical coding discussed in the following section.

#### 4.2.3.4 Coding

While such time-aligned annotations are important to many later uses of the transcribed materials, they nevertheless fall short of identifying individual words or sentences that were translated from the *Fibel*, or of retrieving all occurrences of particular linguistic variables from the corpus. Although corpus users might identify these features by inspecting concordance lines or reviewing entire transcripts, the basic textual annotations supplied by transcription do not allow for all such information to be retrieved automatically. For this to be possible, additional information is needed to associate particular annotations with corresponding prompt words and sentences in the original *Fibel*, as well as with linguistic variables of interest.

To relate utterances in the transcribed corpus with corresponding items in the *Fibel*, text-based identifiers were added on the ‘coding’ tier for each speaker in each transcript. These markers indicated whether this utterance was a response to a particular sentence (S), word (W), or, less commonly, picture (P) in the *Fibel*, and included the number of the corresponding *Fibel* item. Thus, if an utterance for contributor M11 represented a translation of the target word on page 33 of the *Fibel* (‘hunger’), an annotation containing the text ‘W33’ would be added under that utterance on the tier ‘M11-Coding’. Commentary on a particular item was marked with an initial ‘C’ before the identifier. Thus, if contributor M11 commented on some aspect of W11, the utterance containing this commentary received the annotation ‘CW33’. In cases where commentary or translations extended over multiple, pause-delineated utterances, continuations were indicated by appending ‘/c’ to their identifiers. In example (8), contributor F28 rendered

*Fibel* sentence S05 (‘Our boss says that we can have dinner early today’) as three separate utterances, all of which were associated with a single response with the markers ‘S05’ (on the first item) and ‘S05/c’ (on all subsequent utterances that are part of the same response):

- (8) Ons Bauss saigt, daut wi kjennen [...] ons Meddag // tiedig ha'n // von'öag.  
 our boss says that we can our dinner early have: INF today  
**S05** **S05/c** **S05/c**  
 ‘Our boss says that we can have dinner early today.’ (*Fibel* target sentence S05)  
 (F28, 2011-10-27, 3m47s870–3m52s510)

This approach provided a straightforward means of associating translations and commentary in the transcribed corpus with corresponding items in the *Fibel*, even when responses stretched across multiple utterances. This initial coding provided a means of identifying and automatically extracting all responses to a given prompt word or sentence, and represented an important step towards accessing such information consistently in later uses of the corpus. This same scheme was extended to allow for coding of the linguistic variables identified in Section 4.2.1.

Variables’ realizations in a particular utterance were coded as name-value pairs, with the variable names taken from the identifiers listed in Section 4.2.1 and Appendix C (e.g., LXGREY, LXWHETHER, etc.), and their values given in double-quotes after a colon (e.g., lxWhether: “aus”). These name-value pairs were included after the target item identifiers as a comma-separated list enclosed by curly brackets.<sup>51</sup> In example (9), contributor F28’s translation of *Fibel* target sentence S07 is coded as containing instances of four variables: LXSHOULD2S (with the value ‘su’st’), LXMASCACCTHE (with the value ‘dän’), LXMASCACCTHEBIG (with the value ‘grooten’), and LXMASCACCTHEBROWN (with the value ‘bruunen’).

- (9) Du su'st dän grooten, bruunen Boa toch loten.  
 you.SG should the.ACC big.ACC brown.ACC bear leave.alone:INF  
**S07: { lxShould2S: "su'st", lxMascAccThe: "dän",**  
**lxMascAccDefBig: "grooten", lxMascAccDefBrown: "bruunen" }**  
 ‘You should leave the big, brown bear alone.’ (*Fibel* sentence S07)  
 (F28, 2011-10-27, 4m45s890–4m48s490)

<sup>51</sup> This method of representing linguistic variables in corpus coding intentionally resembled JSON, a standard method for representing common data structures in a succinct, human-readable format. As a result, it was possible to use modified JSON parsing procedures to assist in validating the structure of such annotations, ensuring that all such coding was consistently formatted and error-free.

By extending the coding schema used for target items and sentences to linguistic variables, manual coding could be added for attestations of linguistic variation throughout the *Fibel* Corpus. Even with such conventions in place, however, reviewing and adding such codes one-by-one for well over one hundred linguistic variables, spread across 55 target words and sentences from more than 40 contributors in 27 hours of audio, posed a daunting task. While ELAN provided multi-file regular expression-based search capabilities that could identify relevant variants, it did not offer any facilities for batch editing these search results. Thus, for each instance of a linguistic variable successfully retrieved through corpus searches, it was necessary to load the corresponding transcript, add the required coding to the matching utterance, save that transcript, and repeat this process for each of the thousands of items thus identified.

Rather than attempt this coding in ELAN using the above procedure, two small programs, `extract.py` and `reinsert.py`, were developed in Python to assist in this process. The former script allowed regular expression searches to be conducted over a given collection of XML-based transcripts, optionally restricting the search results to a matching set of tiers, data types, and/or contributor identifiers. The results of these searches were returned as a spreadsheet in comma-separated value (CSV) format, with separate columns for each tier in the matching tier hierarchy and, optionally, a set of audio clips extracted from the original corpus recordings for the matching annotations. These spreadsheets of search results and their corresponding audio clips could then be reviewed using common spreadsheet software such as Excel or Open Office Calc, and coding added in bulk to concordance lines in these documents. The latter script then reintegrated all coding in these spreadsheets into the corresponding tiers of the original ELAN transcripts. These twin programs dramatically improved the rate of progress in coding linguistic variables in the corpus, making it feasible to consider the number of variables identified in this study while maintaining the final, time-aligned corpus in an open, XML-based format.

While these coding procedures were well suited to the corpus in its present form, further linguistic annotation of the contents of the corpus (e.g., through lemmatization or part-of-speech tagging) may present another means of identifying and retrieving instances of orthographically discernible variables, and thus facilitate further analysis. Given the relatively high technical requirements for applying such procedures to minority-language corpora (e.g., for lemmatization,

requiring either a computer-readable dictionary or customized program that is capable of providing lemmata for observed word forms; or, for part-of-speech tagging, requiring considerable effort to be invested in developing, training, applying, and reviewing the results of a custom part-of-speech tagging system for the language in question; cf. Cox 2010), these potential alternatives were reserved for future stages of corpus development.

#### 4.2.3.5 Summary

The preceding sections have described the development of the *Fibel* outlined in Section 4.2.2 into a specialized spoken corpus of Saskatchewan Plautdietsch. Proceeding from the initial identification of contributors from both the Mennonite communities of the Saskatchewan Valley and the Catholic communities of St. Peter's Colony through existing contacts and 'friend-of-a-friend'-style introductions (§4.2.3.1), over 27 hours of translations of the printed *Fibel* into local varieties of Plautdietsch were recorded with more than 40 speakers (§4.2.3.2). The resulting audio recordings were subsequently annotated with time-aligned transcriptions in a standardized Plautdietsch orthography (§4.2.3.3), to which further coding was added to associate individual utterances with both target words and sentences in the *Fibel* and with linguistic variables of interest (§4.2.3.4). The final result of these procedures was a fully time-aligned, orthographically transcribed and linguistically annotated spoken corpus of Saskatchewan Mennonite and Catholic varieties of Plautdietsch, stored in open, non-proprietary digital formats that permit a range of later uses.

Although it was thus possible to develop the *Fibel* into this kind of resource, the resulting corpus still has several notable limitations. While the *Fibel* Corpus offers a considerable range of information on variation across Plautdietsch varieties in a relatively consistent fashion, and ensures that variables of interest are well attested across contributing speakers, the controlled nature of the underlying linguistic task represents a shortcoming of this resource, inasmuch as it calls into question the naturalness of contributors' responses. While the provision of accompanying interviews, conversations, or other, less controlled forms of speech fell outside of the immediate scope of this project, such information would nevertheless be beneficial in determining what, if any, influence this translation task may have exerted on the range of

linguistic features considered here.<sup>52</sup> Likewise, observations of apparent convergence between paired respondents merit further attention in continued research. Although these limitations necessarily call for circumspection when interpreting the results of analyses based on this resource, the *Fibel* Corpus in its present form nevertheless provides a unique source of information on variation in Saskatchewan Plautdietsch, and is likely sufficient for the multivariate analysis of variation undertaken in the following chapter.

---

52 Continued collaboration on Plautdietsch-language resource development might aim to include a semi-controlled linguistic task, such as the narration of a wordless children's book or film (cf. Lüpke 2009), as a form of counterbalance against other, more regimented linguistic tasks, although this would necessarily require discussion with partner communities.

## 5 Analysis

### 5.1 Introduction

The preceding chapter gave attention primarily to the methodological and ethical issues raised by the development of linguistic documentation for varieties of Plautdietsch spoken in central Saskatchewan, concentrating on the creation of a specialized, documentation-based corpus capable of capturing a wide range of the variation attested in these communities. With the development of this resource complete, it now remains to be demonstrated that such a corpus is useful not only for local language maintenance and revitalization programs, but also for linguistic analysis. This chapter therefore turns its attention to the specific analytical problems posed by linguistic variation in these communities, drawing on the *Fibel* Corpus to come to a clearer understanding of the linguistic and sociodemographic faces of variation in Saskatchewan Valley Plautdietsch.

At the outset of the preceding chapter, four questions were introduced that frame the present research:

1. What is the extent of linguistic variation in present-day Saskatchewan Valley Plautdietsch? How might this variation be represented to offer insight into the present linguistic situation and to inform further documentation efforts?
2. What correlations, if any, does linguistic variation in the Saskatchewan Valley enter into with the demographic and historical characteristics of these communities and their speakers?
3. Is it possible to identify coherent and recurring patterns of linguistic variation in the Saskatchewan Valley that may suggest distinct varieties? Moreover, is it possible to arrive at such patterns inductively, rather than imposing them top-down through the selection of a limited set of features presumed to be of linguistic interest?
4. How might linguistic research undertaken in these circumstances be consonant with the interests of local communities and result in outcomes that are of benefit not only to linguistic analysis, but also to local language education, maintenance, and revitalization efforts?

The last of these points, which concerns the conduct of language documentation, has partly been addressed in the corpus development described in the preceding sections, but also depends on the

results of the analyses presented in this chapter. The patterning of linguistic variation that the analyses in this chapter seek to describe has implications not only for the documentation and description of this variation as such, but also for the general empirical and practical adequacy of language documentation in communities such as these. Inasmuch as language documentation aims to adequately represent the linguistic practices of a speech community at a given point in time, understanding the nature of linguistic divisions in this community and their relation to subgroups within the community necessarily informs the direction of documentary activities to ensure that relevant linguistic phenomena and their social correlates are not overlooked. This interrelationship between analysis and the prospects and processes of documentation deserves further consideration, and will be taken up in the following chapter.

The remaining three questions focus on a specific range of problems that the linguistic and sociodemographic situation of the Saskatchewan Valley poses for linguistic analysis. First, and perhaps most basic, are descriptive problems that concentrate on assessing the range of variation attested in these communities and its implications for ongoing documentation (Question 1). The following sections therefore aim to give a sense of the extent of linguistic variation in the Saskatchewan Valley (i.e., what features vary between speakers and which forms this variation takes), and to bring attention to areas where further documentation may be needed. These fundamental descriptive problems also relate to the historical and sociodemographic characteristics of these communities (Question 2). Consequently, the following analyses aim to account not only for the linguistic substance of the observed variation, but also its embedding in a particular social and historical context. Finally, there remains the patterning of these linguistic, sociodemographic, and historical features on a larger scale, understanding their distribution across the entire speech community and the divisions they suggest (Question 3). The following analyses must therefore explore these larger patterns, as well, alongside the distributions of individual features that inform them.

Given the volume of the available data and the apparent complexity of the linguistic situation in the Saskatchewan Valley, the following sections undertake several quantitative statistical analyses to address the questions raised above. Section 5.2 approaches the distribution of variation in the *Fibel* Corpus data first from a geographical angle, considering the relevance of physical geography to the patterning of variables representing several areas of linguistic



organization through the use of methods from dialectometry. Applying these methods to linguistic variables, both individually and in aggregate form, representing several aspects of linguistic organization in the corpus provides converging evidence on the patterning of linguistic variation in the region, with each such application profiling different aspects of this variation that might otherwise be missed. Section 5.3 brings together these disparate threads of analysis, drawing on the results of the preceding section to arrive at a fuller profile of the linguistic situation in the Saskatchewan Valley.

## 5.2 *Dialectometric analysis*

Dialectometric methods present a reasonable point of departure into the variation attested in the Saskatchewan Valley, offering several features relevant to the basic descriptive problems raised above. Of particular significance to a first foray into unknown patterns of variation is the facility with which dialectometric methods are able to visualize the distribution of linguistic features geographically. Such visualizations allow possible geolinguistic patterns in variation to be more readily identified, whether at the level of individual linguistic features (e.g., by plotting the geographical distribution of certain variables of interest, or even particular variants of those variables) or on an aggregate level (e.g., by computing overall measures of linguistic distance across a set of variables and considering the resulting relationship between geographical distance and linguistic similarity). Since both these visualizations and their associated distance measures are computationally derived, and thus not limited to small numbers of variables, dialectometric methods are also well suited to the analysis of highly multivariate data, as in the present case, and to drawing such information from existing linguistic resources, such as dialect atlases and corpora (cf. Szmrecsanyi 2011). All of these features are in line with the stated goals of analysis, and will be relied upon below for a preliminary exploration of the *Fibel* Corpus data.

All of the dialectometric analyses presented in this section were conducted using Gabmap, an open-source, web-based application for dialectometric analysis (Nerbonne et al. 2011). Gabmap provides facilities for summarizing, visualizing, and performing statistical analysis on data pertaining to linguistic variation, and offers an accessible interface to a large suite of dialectometric tools that otherwise require considerable expertise to deploy. For its part, Gabmap only requires users to provide a dataset (generally, a tab-delimited spreadsheet

containing the variables of interest and their attested forms at some number of geographically defined 'sites') and a map (in Keyhole Markup Language (KML) format, providing the boundaries of the region of interest and place markers for each site referred to in the dataset). To begin analyzing data from the *Fibel* Corpus in Gabmap, an initial map of the Saskatchewan Valley region was defined in Google Earth, using Dominion Land Survey coordinates from the SaskGrid Township Fabric Map to accurately reproduce the original land allotment boundaries of the Hague-Osler Mennonite Reserve.<sup>53</sup> The Mennonite and non-Mennonite settlements listed in Appendices A and B were then added as additional reference points, alongside several other major geographical features of the region (e.g., the North and South Saskatchewan Rivers, major railway lines, and Highways 11, 12, and 16). This map was then exported in KML format and imported into Gabmap. Figure 5 shows the major boundaries present in this map, including the limits of the historical Hague-Osler Mennonite Reserve.

---

<sup>53</sup> The SaskGrid Township Fabric Map is a publicly available geographical data set that gives the boundaries of the official Saskatchewan land parcel identification system of townships, ranges, and sections in several common GIS formats; see <https://www.geosask.ca/> for more information.

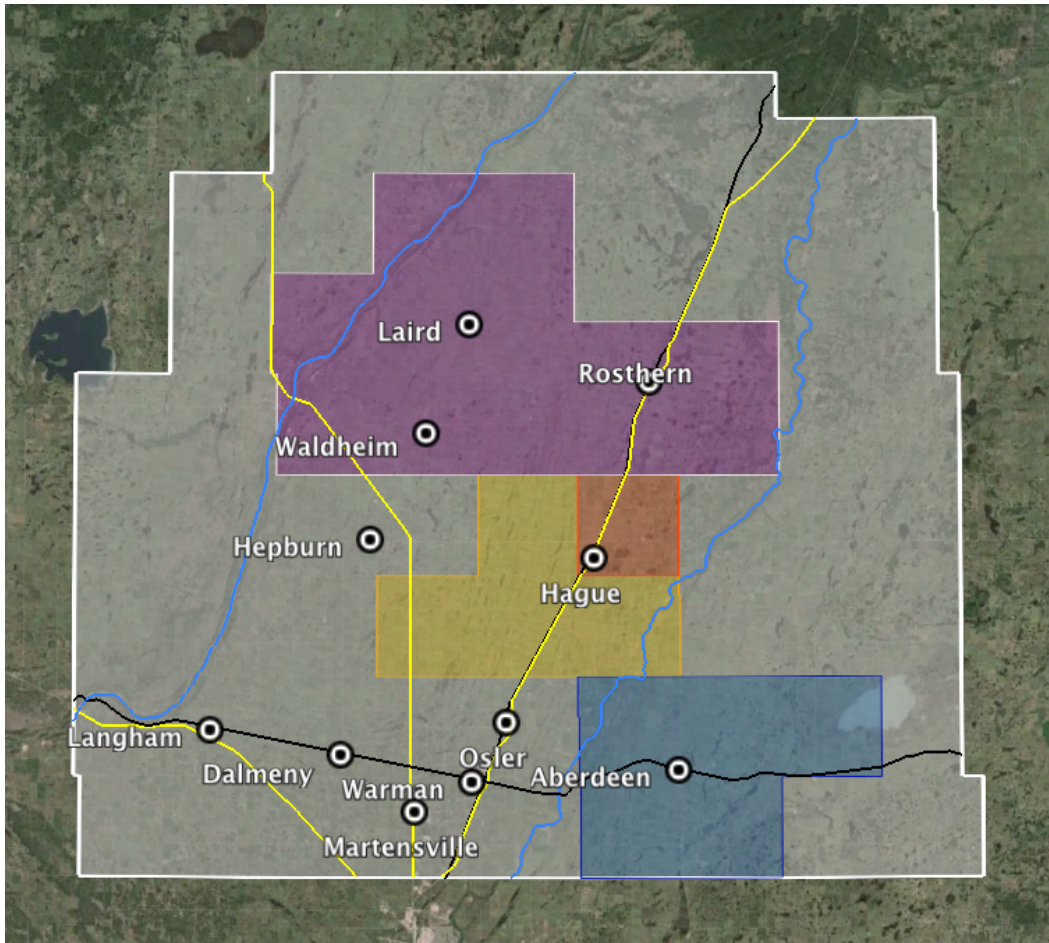


Figure 5. KML layer of the Saskatchewan Valley in Google Earth. Inset coloured areas show the historical boundaries of the Hague-Osler Mennonite Reserve, with the original 1895 allotment in yellow, and later additions shown in orange (early 1898), blue (August 1898), and purple (October 1898).

While defining an appropriate map for use in dialectometric visualization was thus not difficult, the process of importing instances of variation from coding in the *Fibel* Corpus into Gabmap required somewhat more effort. Using the extraction script described in the preceding chapter, it was possible to retrieve all instances of variable coding from the ELAN documents in the *Fibel* Corpus automatically, storing the results in a single, comma-separated spreadsheet. Another short script, `gabmapify.py`, was developed to combine the extracted concordance lines with metadata on each variable and contributor, and to summarize the distribution of variants across contributors in the tab-delimited spreadsheet format required by Gabmap. A third script, `redefine.py`, repeated this aggregation process for the more abstract macro-variables described

in Section 4.2.1, saving the results in the same, Gabmap-compatible format. This three-stage process of extraction, conversion, and redefinition of coded variables was assembled into a single script that converted the entire *Fibel* Corpus data into Gabmap format whenever required. The resulting automated workflow, summarized in (10), thus allowed the *Fibel* Corpus to be further developed in the course of analysis (e.g., to correct transcription errors identified in Gabmap), with any improvements to the underlying corpus becoming immediately accessible to dialectometric analysis without requiring manual correction of data that had been converted to Gabmap format earlier.

- (10)
1. Extract all coding from *Fibel* Corpus in CSV format (`extract.py`)
  2. Convert coding and metadata to Gabmap format (`gabmapify.py`)
  3. Gather results into meta-variables in Gabmap format (`redefine.py`)

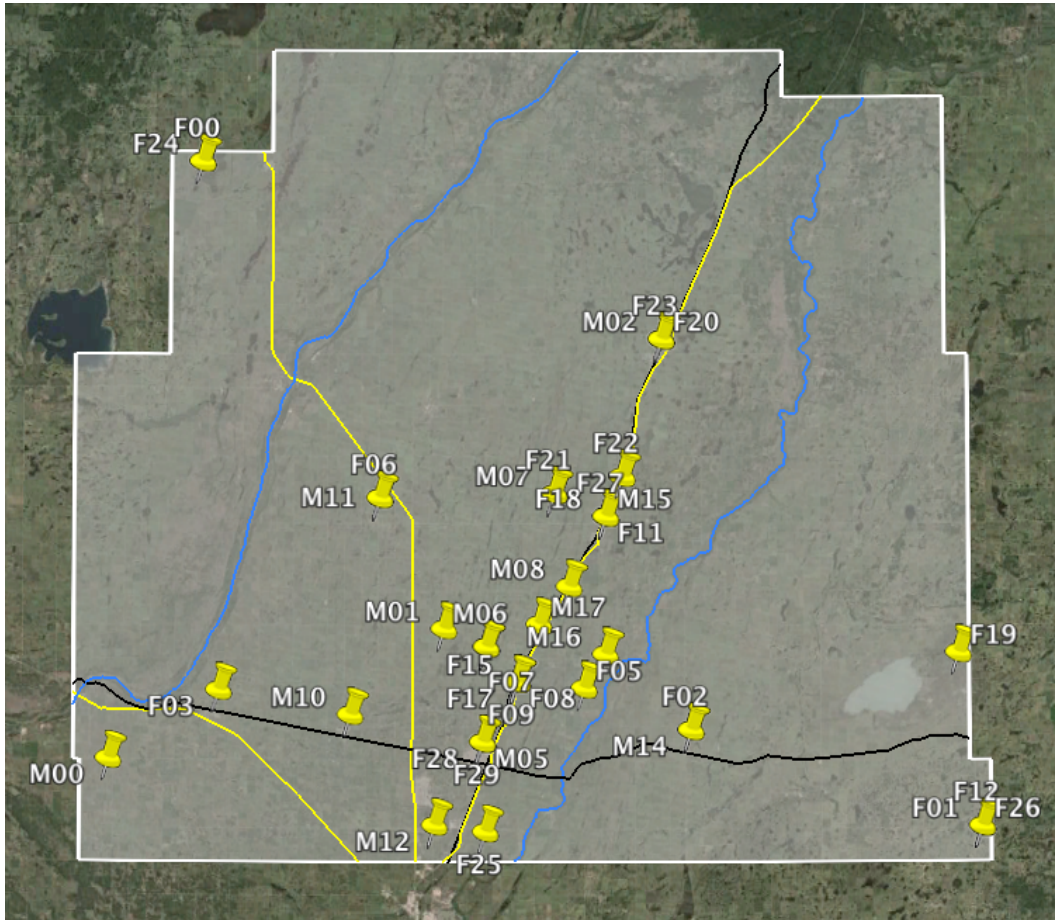
In most dialectometric analyses, information on linguistic variation is typically associated with some number of geographical sites. The relationship between these locations and the sets of linguistic features associated with them is most often the focus of dialectometric investigation. In the present study, however, the primary questions of interest have less to do with linguistic variation that make reference to the properties of *locations* in the Saskatchewan Valley as explanatory variables in their own right, and more with the attributes of individual *speakers* whose lives and linguistic practices are associated in some substantive way with these places. Moreover, with relatively few settlements in the Saskatchewan Valley that could serve as geographical sites in a ‘traditional’ dialectometric analysis (especially given the substantial depopulation of most traditional Mennonite villages in the wake of the forced closure of Mennonite schools and subsequent mass migrations to Latin America; cf. §2.6), and with multiple contributors to the *Fibel* Corpus often available as representatives for the same location, it was decided to visualize geolinguistic variation with reference to speakers, rather than to settlements in the Saskatchewan Valley.

In practice, this decision required an additional stage of conversion to be added to the workflow in (10) to create reference points on the map that refer to speakers, rather than settlements. Using a small, custom script, `speakermap.py`, it was possible to automatically

remove all existing site markers from the KML map layer (e.g., “Warman”, “Blumenheim”, etc.), replacing them instead with markers at each corpus contributor’s place of birth (e.g., setting a marker with the label “F05” at the location of Kronsthal, where contributor F05 was born and raised).<sup>54</sup> Although this represented a departure from standard dialectometric practice, shifting the focus from single representatives of a large number of sites to many representatives of a smaller number of sites, this did not pose any significant problems for Gabmap, for the above workflow, or for the larger aim of giving consideration to the possible explanatory role of physical geography in the distribution of variation among speakers in the corpus. Figure 6 shows the automatically amended KML map of the Saskatchewan Valley, with the birth-places of all contributors to the *Fibel* Corpus displayed as appropriately labelled place markers.

---

54 For speakers whose places of birth fell outside of the Saskatchewan Valley (e.g., the Catholic Plautdietsch contributors from the Humboldt area), a fixed location was chosen at the edge of the regional boundaries that was closest to their actual location. Given the potential for such locations to be accidentally misinterpreted as sites in the Saskatchewan Valley, their marked placement at the extreme edges of the map boundaries has been supplemented with an explicit listing of all such places and speakers in Table 17.



Dialectometric methods offer considerable freedom in exploring variation in these corpus data. While these approaches are able to aggregate and analyze all available sources of information on linguistic variation at once, it is sometimes also worthwhile to consider the patterning of variables representing different aspects of linguistic organization on a more detailed level (cf. Spruit 2008: 65–90). An aggregate analysis may bring larger-scale trends to light that would otherwise go unnoticed, but is perhaps less well suited to answering questions about the relationships between different kinds of variables (e.g., are trends in variation observed with sounds and sound patterns also reflected in the same way in variation related to lexical selection, word formation, or other aspects of linguistic structure?). To address questions such as these, the following sections first attend to patterns of variation encountered in phonemic and phonological variables (§5.2.1), lexical variables (§5.2.2), morphological variables (§5.2.3), and syntactic variables (§5.2.4), before returning to examine the results of aggregate analysis over all of the available features (§5.2.5) and considering the implications of these results for our general

understanding of linguistic variation in the Saskatchewan Valley (§5.3).

### 5.2.1 Phonological clustering

The above process of extracting coded instances of linguistic variation from the *Fibel* Corpus resulted in no shortage of information on phonemic and phonological variability, producing some 3,618 coded instances of the 85 sound-related items represented in the corpus. This represents much more data than could realistically be treated by reviewing the data ‘by hand’. At a basic level, dialectometric tools such as Gabmap allow one to quickly summarize contributors’ responses for individual variables and to visualize their geographical distributions. Figure 7 shows the distribution of  $v_{\text{CANPLVOWEL}}$  across the Saskatchewan Valley. As is apparent from these plots, the more common of these variants is *kjenne(n)* ‘can (pl.)’, which occurs particularly densely in the central area of the Saskatchewan Valley. The other variant, *kjänne(n)*, is less commonly attested, generally appearing more towards the periphery of the region and in speakers from outside of the Saskatchewan Valley proper (e.g., the three Humboldt contributors in the cluster shown in the lower right-hand corner of the plot). Since this variable has only two attested variants, the two plots in Figure 7 are mirror images of one another (i.e., where one map has no attestations of a form, shown in white, the other has the remaining attestations, shown in blue), each summarizing the complete distribution of this variable across the region.

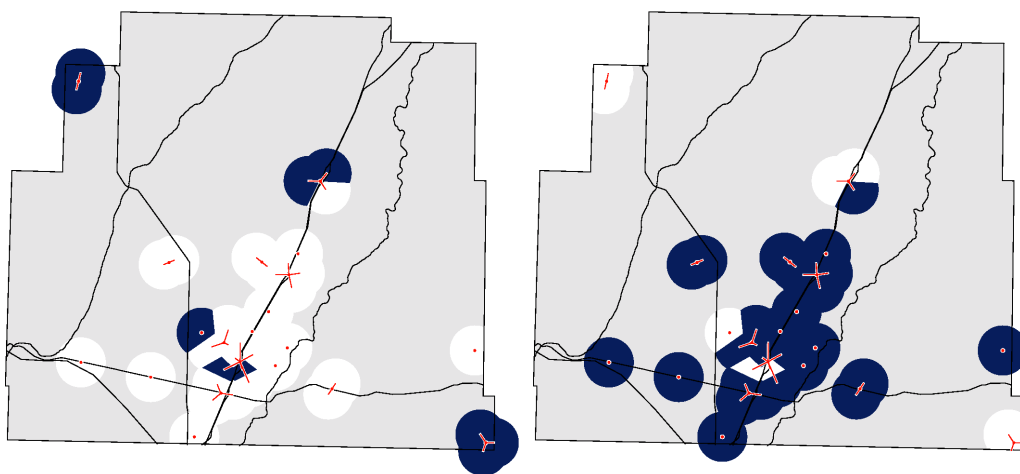


Figure 7. Distribution of  $v_{\text{CANPLVOWEL}}$ , with [e] (*kjänne(n)*) on left and [ɛ] (*kjenne(n)*) on right.

Also apparent from these visualizations is the perfect regularity of the division between contributors with respect to this variable: no speakers are attested as using *kjenne(n)* in some cases and *kjänne(n)* in others. This is not the case with many other variables. In Figure 8, while most contributors show no variation in their selection of either fronted or non-fronted realizations of the *oa* diphthong in pre-velar environments (e.g., *Doag* vs. *Döag* ‘days’), several speakers (e.g., M00, shown on the far left of the plots) are attested as using both forms. These mixed realizations are shown in cyan in the distribution plots.

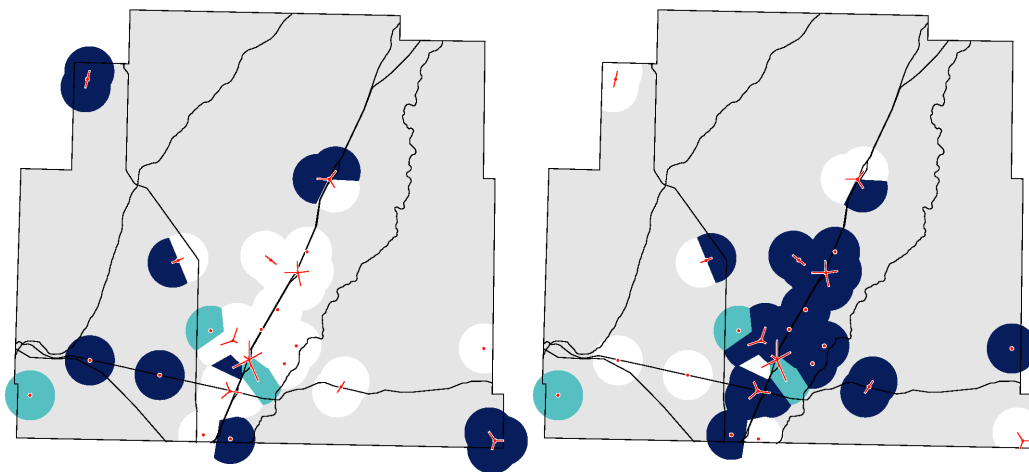


Figure 8. Distribution of  $v_{\text{REALIZATIONOAPREVELAR}}$ , with non-fronted realizations (e.g., [oə], [oː]) on left and fronted realizations (e.g., [øø], [eø]) on right.

In the above distribution plots, there again appears to be a division between the central Saskatchewan Valley, where fronted realizations of  $v_{\text{REALIZATIONOAPREVELAR}}$  are most common, and the peripheral region (i.e., both north and south of the core Mennonite settlement area, as well as west of Highway 12, which vertically demarcates the westernmost third of the region). This pattern is reflected in the distribution of the phoneme /u/ summarized in Figure 9, as well, where back realizations are more common in the periphery of the valley, and front realizations in the centre, albeit with somewhat more variability than in the preceding example.



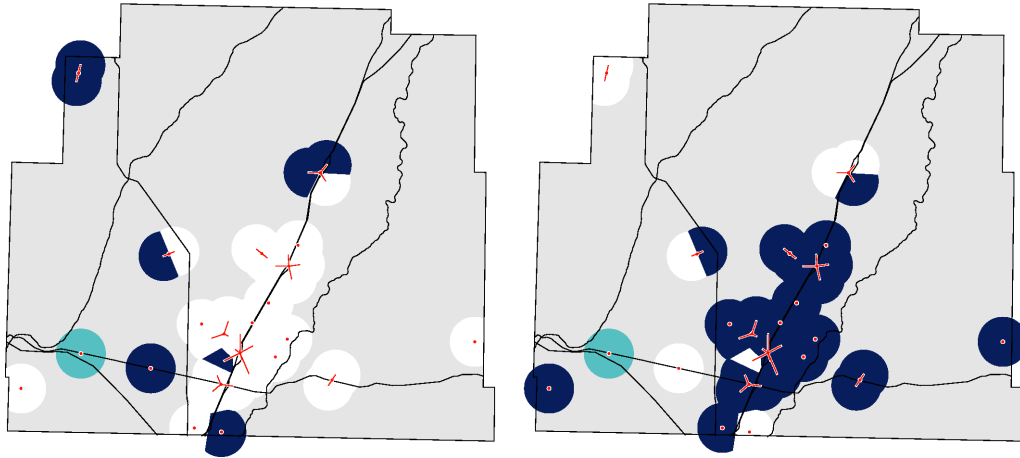


Figure 9. Distribution of  $vU_E$ , with back realizations (i.e., [u]) on left and front realizations (i.e., [y]) on right.

Other distribution plots suggest marked linguistic differences between the Catholic and Mennonite contributors. In Figure 10, phonemic variables associated with *eea* (as in *vea* ‘four’) and *oo* (as in *Foot* ‘foot’) are shown to divide these two groups of contributors evenly: where the Catholic contributors have [ɐɐ(r)] and [oː] for *eea* and *oo*, Mennonite contributors have [əɐ] and [əʊ], respectively. As we will see below, these are not the only differences between speakers from these two groups, and other dialectometric techniques give us several means of further investigating the linguistic relationship between Catholic and Mennonite Plautdietsch speakers in Saskatchewan.

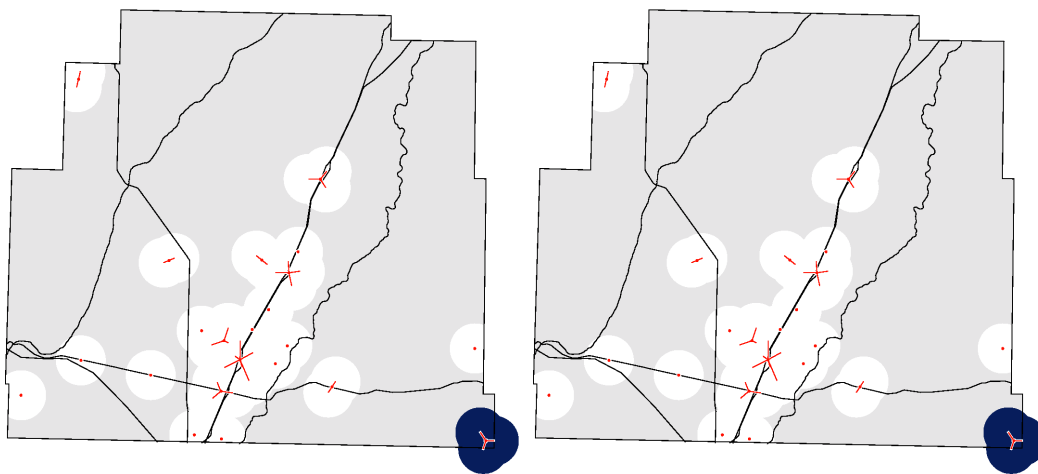
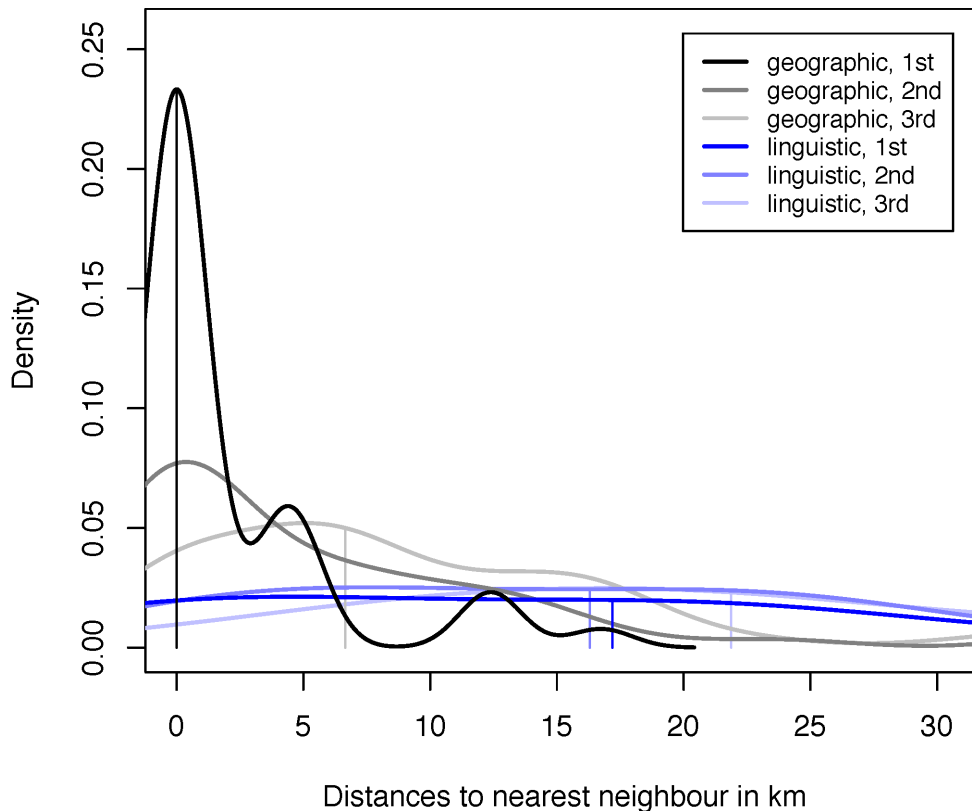


Figure 10. Distribution of  $vEEA$  (left, [ɐɐ(r)] in blue) and  $vO_o$  (right, [oː] in blue).

Simple visualizations such as these can be quite helpful when formulating initial hypotheses about the distribution of variation. The distributions seen above suggest several possible dimensions of linguistic differentiation between groups in the area, separating both Catholic and Mennonite and central and peripheral groups. While it is possible to consider each of the 85 sound-related variables like this in turn, comparing their geographical distributions to identify patterns of differentiation, dialectometric tools such as Gabmap have more to offer than just the visualization of particular variables of interest. As noted above, one of the significant strengths of dialectometric methods lies in their ability to present aggregate views of variation across a large number of variables, without privileging any particular subset of variables in advance. These higher-level abstractions provide valuable counterbalance against selective presentation of variables—it may be that the above distribution plots are not representative of the overall pattern of variation encountered in the Saskatchewan Valley, or are potentially confounded by other countervailing trends in the data—and thus offer another means of ensuring due circumspection in the reported divisions.

As one of its most basic aggregation functions, Gabmap is able to summarize the distance between each contributor and his or her nearest neighbour, whether defined geographically (i.e., in terms of kilometric distance) or linguistically (i.e., in terms of overall similarity in the selection of variants for a given set of variables). This allows us to consider the relationship between linguistic differentiation and physical geography in more detail. If contributors are linguistically more similar to individuals who live close to them, then we would expect the average distances to both their geographical and linguistic nearest neighbours to be low: those individuals with whom they share the greatest number of linguistic features in common are close at hand. If, on the other hand, contributors with similar patterns of linguistic variation are dispersed throughout the region, then there should be little correlation between the average distance to geographical and linguistic nearest neighbours: the most linguistically similar individuals may live some distance away, and individuals living nearby may demonstrate quite different linguistic features.



*Figure 11.* Density plot of distances to nearest geographical and linguistic neighbours, phonemic-phonological variables.

Figure 11 summarizes the average distance from contributors in the Saskatchewan Valley to their nearest geographical and linguistic neighbours (where the labels ‘1st’, ‘2nd’, and ‘3rd’ refer to the first, second, and third-closest neighbours, whether by linguistic or geographical distance). This density plot suggests a different relationship between geography and linguistic variation than the preceding distribution maps made apparent. From this figure, it is clear that, while speakers are generally clustered quite tightly geographically (hence the low average distance to their nearest geographical neighbour), these nearest neighbours are rarely the individuals with whom they have the greatest phonological similarity. That is, despite relatively dense settlement, there appears to be little relationship between geographical proximity and linguistic similarity with respect to phonological variation: the distance from contributors to their linguistically most similar counterparts varies considerably, resulting in the flat density distribution of distances to

nearest linguistic neighbours seen above. Although it may still be the case that certain regions of the Saskatchewan Valley may pattern similarly with respect to their selection of phonological and phonemic variants, on the whole, phonological variation in the region does not appear to be distributed in accordance with the common dialectological assumption that greater geographical distance generally correlates with greater linguistic dissimilarity. At least for these variables, the situation in the Saskatchewan Valley appears much more varied than one would be led to suspect from the individual distribution plots seen above, and does not resemble a typical, undisturbed dialect continuum. (As this observation is limited to sound-related variables for the moment, we will revisit this point with a wider range of variables in the following sections.)

The preceding summary is thus based on the aggregation of geographical and linguistic distances between all pairs of speakers in the Saskatchewan Valley. While this perspective on the linguistic situation suggests a generally more piecemeal distribution of patterns of variation across the region, it may be the case that this does not hold for all areas equally. For some groups of speakers, it may be that geographical distance and linguistic distance *are* correlated, and that these relationships are simply lost when averaged with all other contributors. This possibility can be investigated further by considering reference-point maps, which present information about linguistic similarity and geographical distance from the perspective of a particular contributor. Figure 12 presents a reference-point map for contributor F01, one of the Humboldt-area Catholic contributors to the *Fibel* Corpus. At least for this group of speakers, the ‘nearest neighbour’ hypothesis appears to hold true: contributor F01 shows the greatest linguistic similarity to other speakers from the Humboldt area, and much less affinity to Mennonite speakers elsewhere (with the possible exception of F23, who appears in light blue in the upper half of the lefthand panel). For this cluster of speakers, geographical and linguistic similarity appear generally correlated, although given the limited number of Humboldt-area speakers included in this sample, it is not possible to test whether this group is as linguistically coherent as they appear, or if this correlation would remain as pronounced when additional speakers and settlement areas from that region were considered.

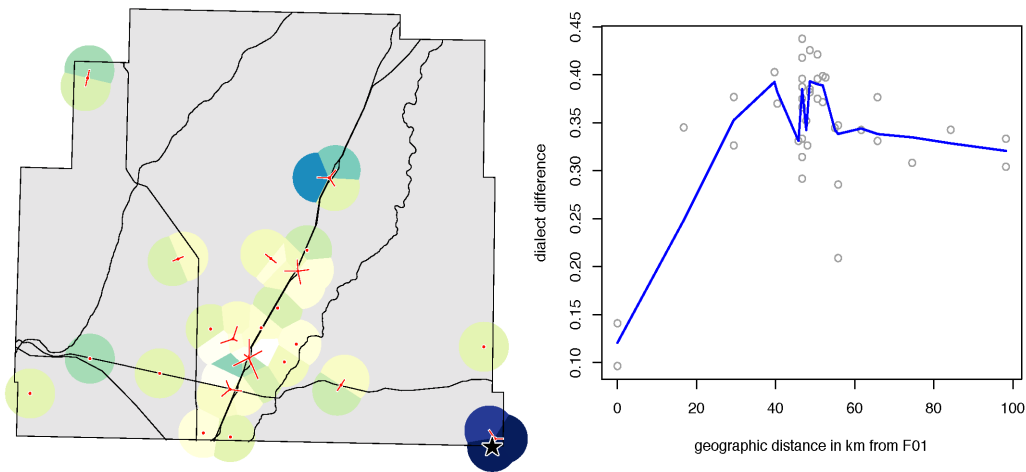


Figure 12. Reference-point map for contributor F01.

Using the same methods, it is also possible to look more closely at the potential division between central and peripheral Saskatchewan Valley Mennonite Plautdietsch speakers that was hinted at in earlier distribution plots. Figure 13 and Figure 14 present reference point maps for contributors F15 and M00, respectively, with the former representing the central region and the latter the peripheral areas of the valley. In both cases, we find both considerable similarity with nearby speakers, accompanied by intervening exceptions. While the linguistically most similar contributor to F15, M04, is from the same settlement, the second-most similar contributor, F11, is from much farther north, with many other, less similar contributors appearing between them. This is even more pronounced with M00, who, while linguistically close to M01 at the western edge of the central cluster, bears the greatest overall similarity to F19 at the opposite end of the geographical spectrum. Reference-point maps such as these suggest a non-trivial relationship between geography and phonological variation in the region, one that is not easily reduced to a single statement that holds absolutely for all speakers or communities. While speakers in the central cluster appear to share more linguistic features in common with one another than do speakers in the peripheral areas, both regions show considerable internal variation in the geographical patterning of linguistic differences, and thus are unlikely to represent coherent dialect continua in the traditional sense.

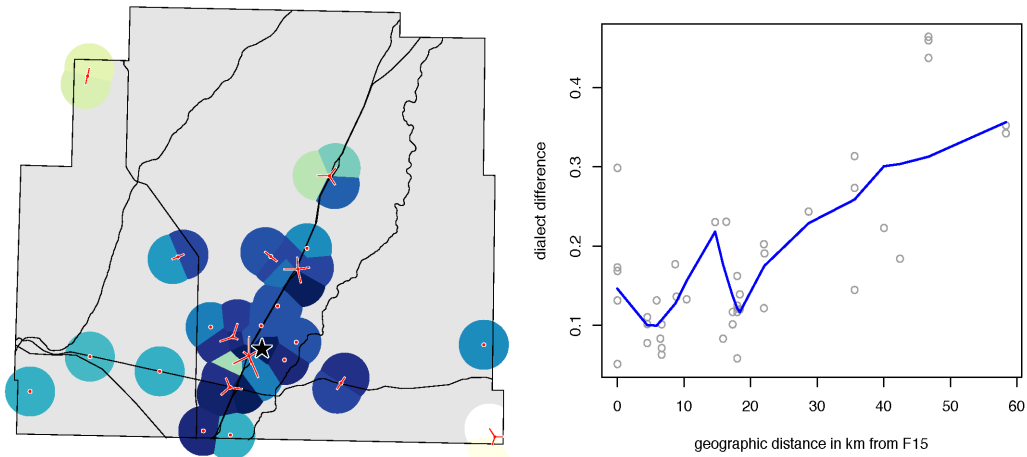


Figure 13. Reference point map for contributor F15.

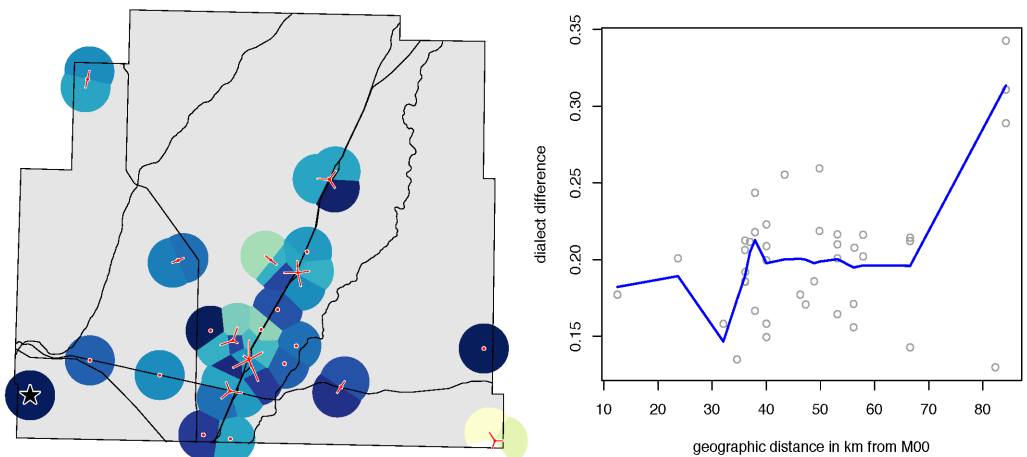


Figure 14. Reference point map for contributor M00.

Reference maps such as these point to significant similarities between Plautdietsch speakers in Saskatchewan, even while demonstrating considerable unevenness in their geographical distribution. These linguistic similarities might also be considered on their own as a means of identifying groups of speakers who share common conventions of variation, whether or not these similarities are reflected geographically. As a first step in this direction, multi-dimensional scaling (MDS) may be performed on the linguistic distances computed by Gabmap for the given set of variables, and thus reduce the relationships between all pairs of speakers for all variables to a lower-dimensional space that can be more easily visualized. Figure 15 shows the results of multi-dimensional scaling as applied to the linguistic distances between contributors for all

phonemic and phonological variables, displaying the outcome in two dimensions.

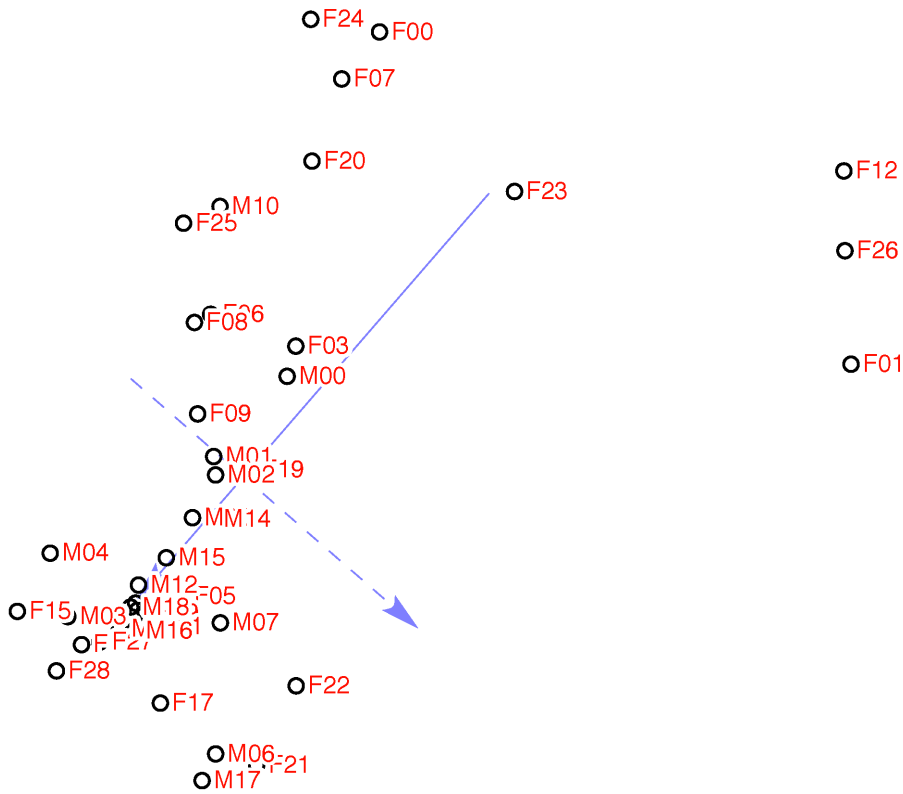


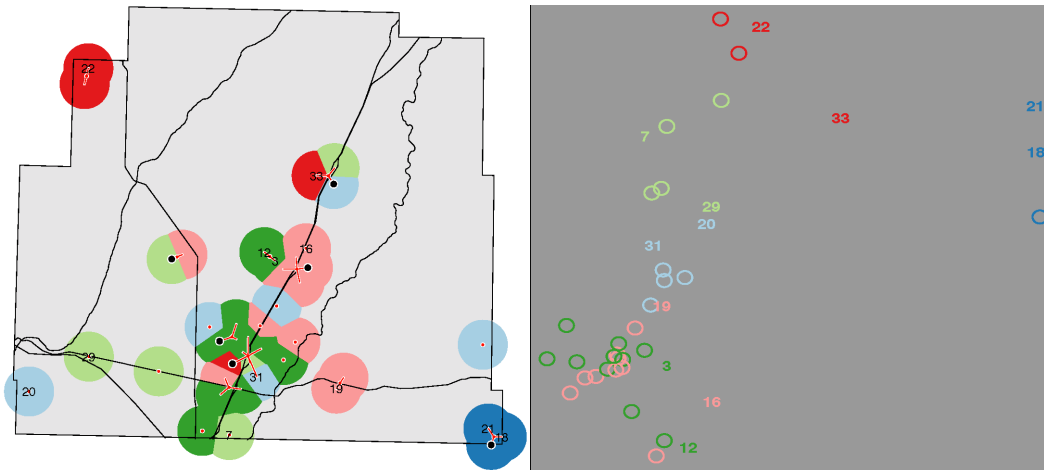
Figure 15. Two-dimensional MDS plot, phonemic-phonological variables ( $r = 0.96$ ).

This representation provides further support for several of the divisions identified above. Most immediately, there is a marked separation between speakers F01, F12, and F26 and the rest of the contributors. These are the three Humboldt Catholic contributors to the *Fibel* Corpus, who have quite different phonological-phonemic patterns from the majority of the Mennonite contributors (although bearing a greater overall similarity to speaker F23 and contributors appearing in the upper half of this plot). Interestingly, the remaining Mennonite speakers pattern more or less consistently along a single axis, with no marked breaks between groups of speakers and somewhat more variability among the lower group of speakers. At least preliminarily, this might be taken to suggest greater continuity between Mennonite varieties of Plautdietsch than would be anticipated from the less consistent geographical distribution observed above, with speakers falling not into wholly disparate clusters, but rather ranging across the phonological and phonemic variants found in the present variables.

Multi-dimensional scaling can also be used in combination with hierarchical clustering methods to help identify groups of speakers who share similar conventions in their use of phonological and phonemic variants. Tools such as Gabmap not only provide implementations of standard agglomerative hierarchical clustering procedures that operate on aggregate feature distances to arrive at groupings of speakers, but, importantly, also allow for such clusters to be evaluated through comparison with a corresponding two-dimensional MDS plot. This affords some constraint on possible clustering analyses, which could otherwise be made as complex or as simple as desired to fit preconceptions of the data. If the proposed clusters correlate with separate ‘clouds’ of points in the MDS plot, then they are likely to represent actual discontinuous feature sets in the data that would be appropriate to represent as clusters. On the other hand, if no clear sets of points are found in the corresponding MDS plot, then the data are likely continuous and further argumentation would be required to justify the proposed groups of speakers.

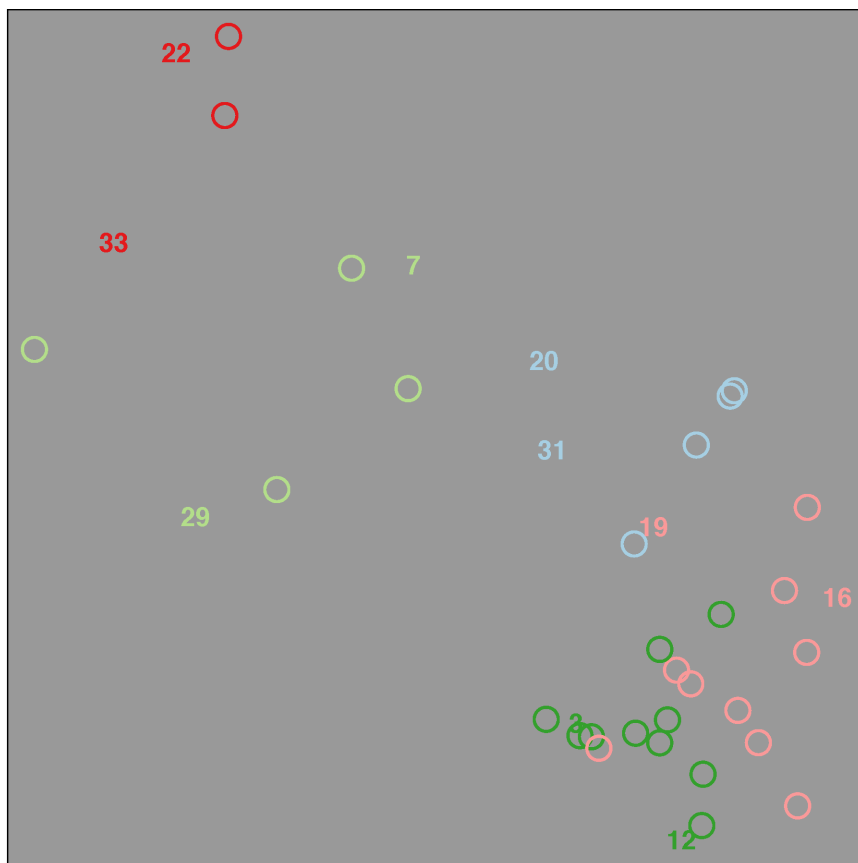
Figure 16 presents one such application of agglomerative hierarchical clustering to the present data, using Ward’s Method to iteratively group speakers into progressively larger clusters that have the minimum possible variance between them. The preliminary six clusters produced by this method are plotted on a map of the Saskatchewan Valley, and are accompanied by a corresponding two-dimensional MDS plot. As in the preceding figure, it is clear that the Catholic speakers in dark blue form a cluster separate from the other, non-Catholic speakers, an observation that is confirmed by the distance between those speakers and all others in the MDS plot below.



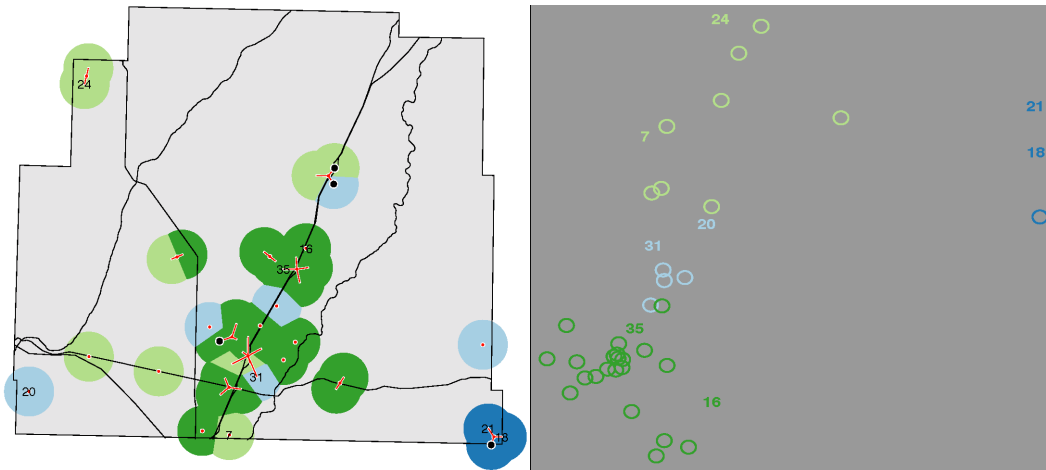


*Figure 16.* Hierarchical clustering by phonemic and phonological features (Ward's method, six clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.96$ ; right).

The remaining five clusters that represent the Mennonite Plautdietsch speakers do not appear to be separate from one another in the MDS plot, raising questions as to the validity of these divisions. Removing the cluster of Catholic speakers from consideration temporarily, we see in Figure 17 a much clearer depiction of the relationship among the remaining groups. Based on this information, it is doubtful that the overlapping clusters shown in dark green and light pink in the lower right-hand corner of the MDS plot represent stable divisions between groups of speakers. Likewise, it is not clear that the two clusters in red and light green in the top-left quadrant are entirely separate, suggesting that a reduction in the overall number of clusters may be appropriate.

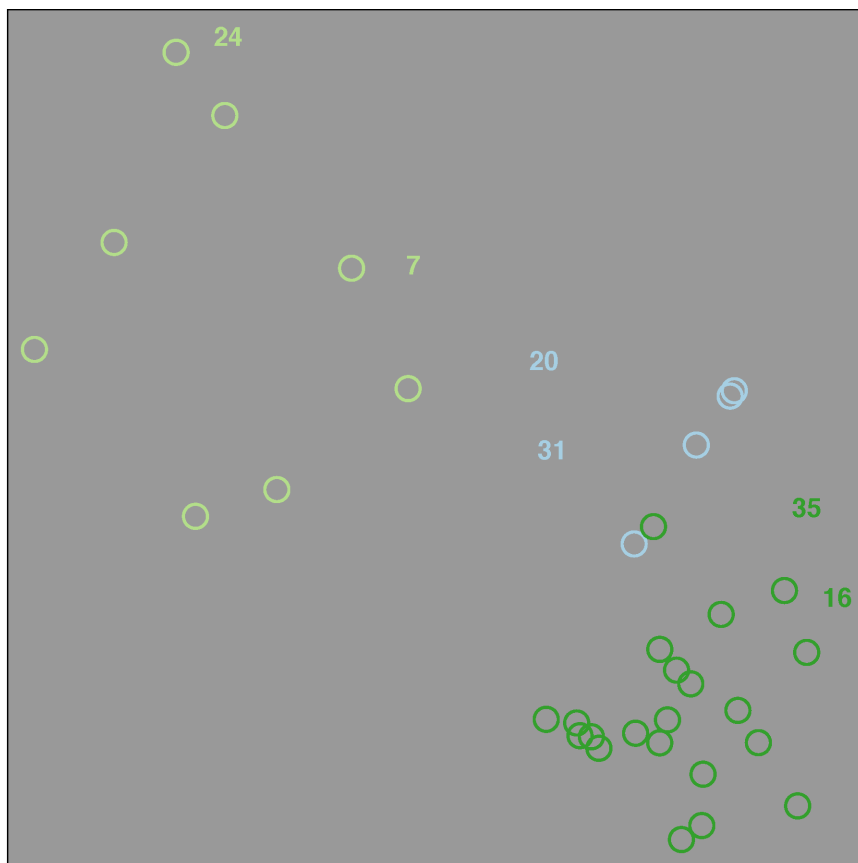


With several of these clusters open to question, the hierarchical clustering analysis can be re-run with a smaller number of target clusters, successively comparing the results of clustering with the corresponding multi-dimensional scaling analysis until arriving at a set of clusters that is supported by the latter results. For the present phonological and phonemic features, correspondence between the clustering and MDS analyses is reached with four clusters, as seen in Figure 18 below.

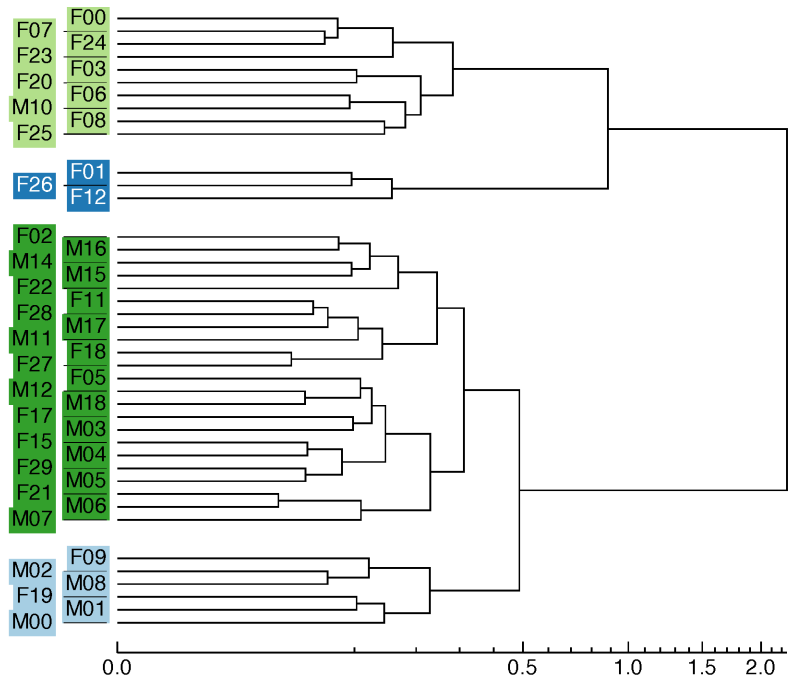


*Figure 18.* Hierarchical clustering by phonemic and phonological features (Ward's method, four clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.96$ ; right).

While the cluster of Catholic speakers remains stable across all of these analyses, the earlier clusters of Mennonite speakers in the top left and bottom right quadrants of the MDS plot have now collapsed, leaving three divisions among those speakers. Temporarily excluding the Catholic speakers from the MDS plot in Figure 19 again provides useful detail on the structure of the Mennonite-only clusters. This reveals two larger, separate clusters in light and dark green, as well as a smaller, transitional cluster between them in light blue that appears somewhat less clearly defined.



Interestingly, the transitional cluster in Figure 19 appears to correlate with those speakers most similar to contributor M00, as seen in Figure 14. For the remaining Mennonite clusters, however, it would be useful to have a clearer sense of the speakers belonging to each group. An alternative visualization of the same clustering analysis, this time in the form of a dendrogram, provides this information in Figure 20. This representation exposes several features of these clusters that were not readily apparent from the earlier figures. From this, we note that the cluster of Catholic speakers are most similar in their phonological and phonemic features to the cluster of Mennonite speakers, while the transitional light blue cluster has a closer linguistic relationship to the dark green cluster at the centre of the Saskatchewan Valley.



Clusters such as these, arising from an aggregate analysis of sound-related features, raise obvious questions as to their basis and motivation: do these linguistically founded groupings have any relationship to the non-linguistic characteristics of the speakers they comprise? While dialectometric tools such as Gabmap permit some exploration of physical distance as an explanatory factor—recall the comparison of geographical distances to nearest linguistic and geographical neighbours seen in Figure 11, for instance, which suggested a more piecemeal distribution of variant usage patterns throughout the Saskatchewan Valley—they do not generally offer additional functions to explore other, non-geographical factors. Nor is this entirely unexpected: dialectometric tools commonly seek to account for linguistic variation by reference to geographical placement, modelling sites, rather than speakers. In the present application of these tools to the Saskatchewan Valley, however, this is not the case, and investigations of possible correlations between these linguistic groupings and other sociodemographic factors must be explored separately. Other analytical methods, such as multiple correspondence analysis (MCA) and generalized additive modelling (GAM), are capable of serving this function, providing complementary information to the dialectometric analyses presented here, although their application falls outside of the scope of this study (see §6.1 for further discussion).

Even with this notable limitation concerning non-geographical predictors of linguistic

variation, dialectometric methods do allow for further analysis to be performed on the clusters themselves, providing information on which linguistic features are most distinctive in each group. In Gabmap, this can be accomplished through several different procedures, including the distance-based cluster determinant method described by Prokić, Çöltekin and Nerbonne (2012). In this approach, items are identified in each cluster of interest that are maximally similar for all cluster members and maximally different from members of other clusters. After normalization, these two distances—the ‘within distance’ (the average difference between members of a cluster for a given item) and the ‘between difference’ (the average difference between members of this cluster and non-members for a given item)—are combined to serve as an aggregate score of the degree to which a given item can be considered distinctive in the cluster of interest. Applying this procedure to the above clusters identifies several such ‘shibboleth’ variables. Table 19 presents the four highest-ranked cluster determinants in each of these groups (scores indicate the overall distinctiveness of the item, followed by its within distance and between distance):

<i>Dark blue</i>	<i>Light blue</i>	<i>Light green</i>	<i>Dark green</i>
vWASVOWEL (2.55, -0.40, 2.52)	vAuEeGAVE (1.02, -1.02, 0.06)	vOEA (1.49, -1.19, 0.90)	vREALIZ.UAPREVELAR (1.19, -0.99, 0.67)
vOo (2.55, -0.40, 2.52)	vNj (0.93, -0.93, -0.19)	vREALIZ.UAPREVELAR (1.26, -0.99, 0.79)	vREALIZ.OAPREVELAR (1.18, -1.00, 0.63)
vHORSESVOWEL (2.55, -0.40, 2.52)	vAUJ (0.81, -0.81, -0.19)	vUEA (1.18, -0.97, 0.67)	vOEA (1.19, -0.99, 0.67)
vEe (2.55, -0.40, 2.52)	vAuEeOoTOOK (1.02, -1.02, 0.06)	vREALIZ.FINALREN (1.15, -1.13, 0.22)	vCOULD2SRREDUCTION (1.00, -0.93, 0.36)

*Table 19.* Distance-based cluster determinants, phonological-phonetic variables.

While these scores provide an indication of the degree of distinctiveness of each variable’s distribution across clusters, they remain relatively opaque in other respects. It is not immediately apparent from these scores which particular variants of each variable are associated with each cluster, or with which other groups they may be shared. Individual distribution plots of these variables shed light on these aspects of their occurrence, particularly when compared against the corresponding plot of the clusters identified in Figure 18. The distributions of the highest-ranked determinants for the cluster in dark blue are shown in Figure 21 below.

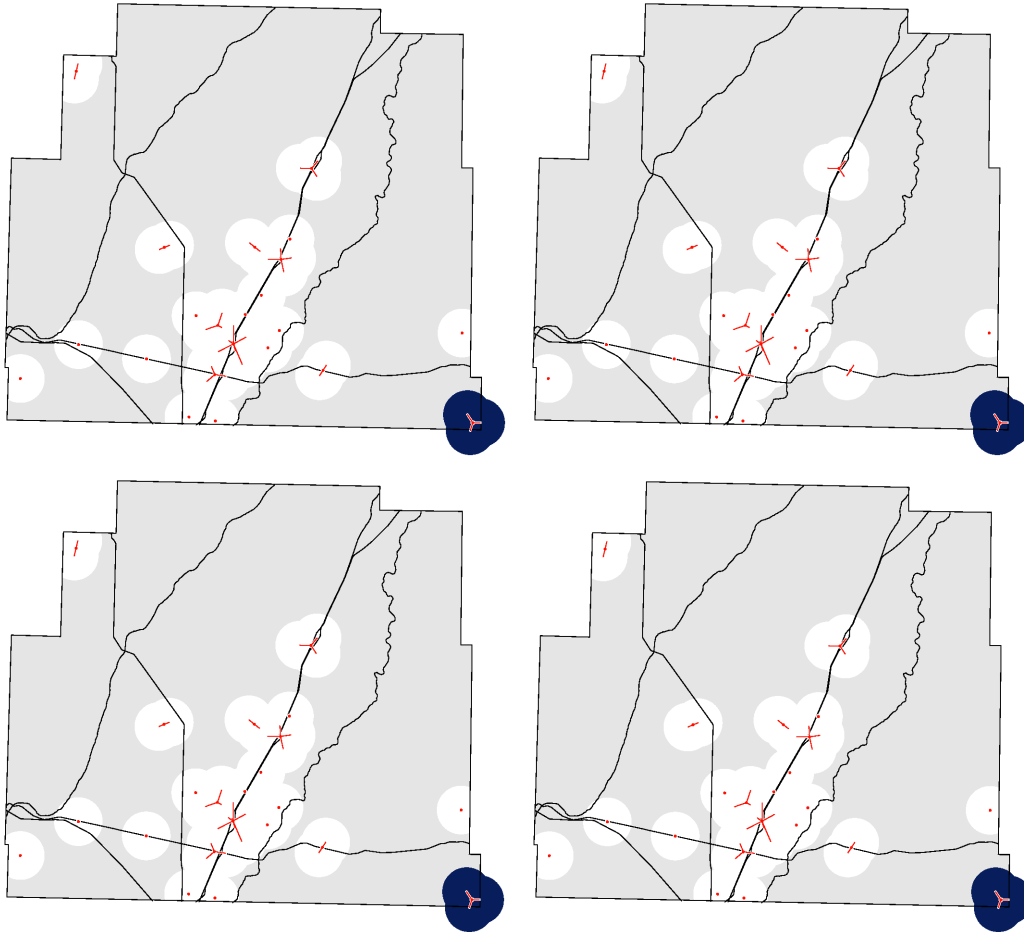
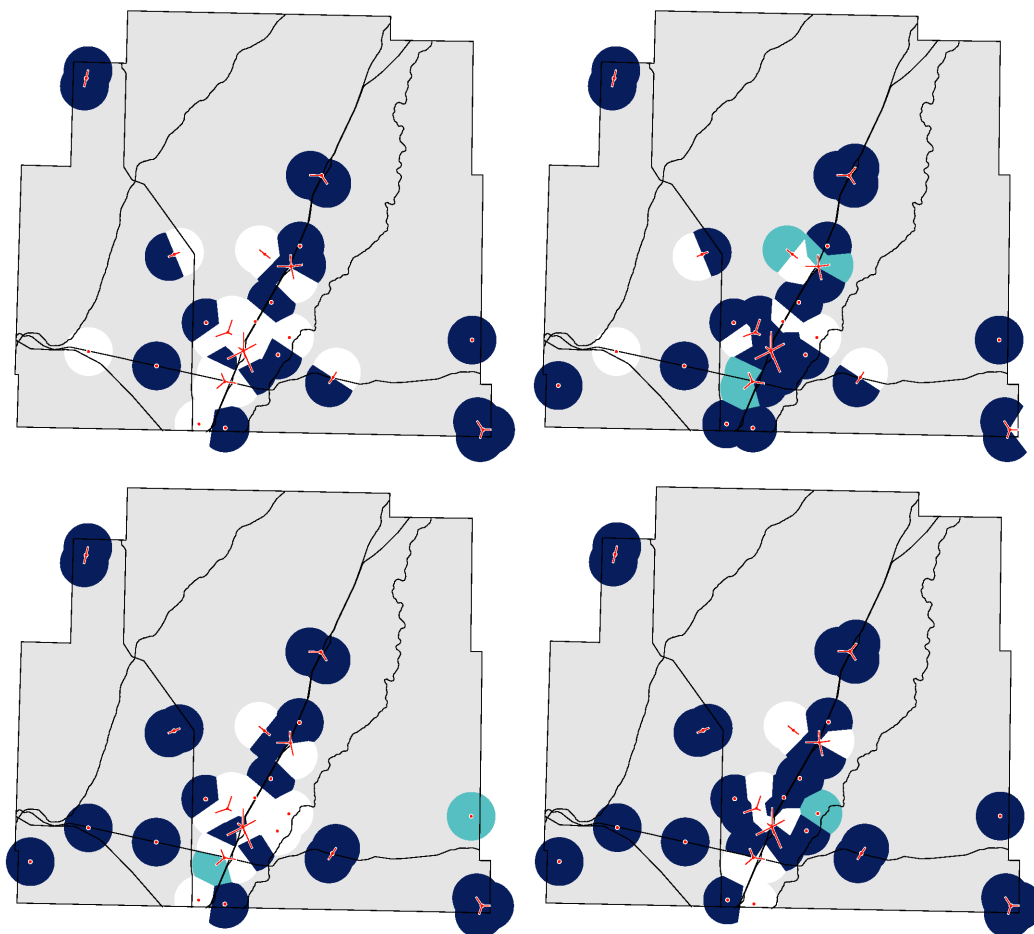


Figure 21. Distribution of phonemic-phonological cluster determinants, dark blue cluster. Clockwise from top left:  $v_{WASVOWEL}$  ([eə] in blue, [iə] elsewhere),  $v_{Oo}$  ([oː] in blue, [əʊ] elsewhere),  $v_{HORSESVOWEL}$  ([eə] in blue, [iə] elsewhere),  $v_{EE}$  ([eː] in blue, [ɛɪ] elsewhere).

It is clear from this visualization that several stark phonological differences separate the Catholic and Mennonite contributors to the *Fibel* Corpus. Where all Mennonite Plautdietsch speakers have [iə] for the phoneme /ea/ in both *wea* ‘was’ and *Pead* ‘horses’, all of the Catholic speakers have [eə]. More strikingly, the diphthongs found in Saskatchewan Mennonite varieties of Plautdietsch for the phonemes *oo* and *ee* are wholly absent from the represented Catholic varieties, where long monophthongs are found instead. It is worth noting that all such differences are maximally distinctive: all such features are shared by each of the Catholic contributors, and are found nowhere among the Mennonite contributors.

By comparison, the phonological differences between the remaining clusters of Mennonite Plautdietsch speakers are less marked (as is apparent from their lower cluster

discriminant scores in Table 19), but nevertheless provide valuable insights into the linguistic patterning of each group. Figure 22 shows the distribution of top cluster determinants for the transitional, light blue group.



*Figure 22.* Distribution of phonemic-phonological cluster determinants, light blue cluster. Clockwise from top left:  $vAuEeGAVE$  (<au> in blue, <ee> elsewhere),  $vNj$  ([ɲ] in blue, [ɲ] elsewhere),  $vAUj$  ([aʊ] in blue, [ɔj] elsewhere),  $vAuEeOoTook$  (<au> in blue, <ee> and <oo> elsewhere).

These distribution plots make the reasons behind the lower discriminant scores for this cluster of speakers clearer: even where all members of this cluster show similar variant preferences (e.g., with  $vNj$ , where all six members of the cluster have [ɲ]), these same variants are also shared with a considerable number of speakers in other clusters. Indeed, none of these features is shared exclusively with members of one of the other clusters: rather, as can be seen, the features that



members of the light blue cluster have in common with other speakers are found throughout the rest of the Saskatchewan Valley. This reflects the low between distances in Table 19: while speakers in this cluster are generally coherent in their selection of variants, these same variants are also found in other clusters—particularly among other Mennonite speakers, as the multi-dimensional scaling analysis in Figure 19 and the dendrogram in Figure 20 suggest.

Whereas the light blue cluster thus shares many of its patterns of variant selection with members of other groups, a greater degree of distinctiveness is found between the remaining two clusters, which contrast notably in several features. Figure 23 summarizes the geographical distribution of the highest-ranked phonemic-phonological determinants of the peripheral, light green cluster, while Figure 24 shows the phonological determinants of the central, dark green cluster.

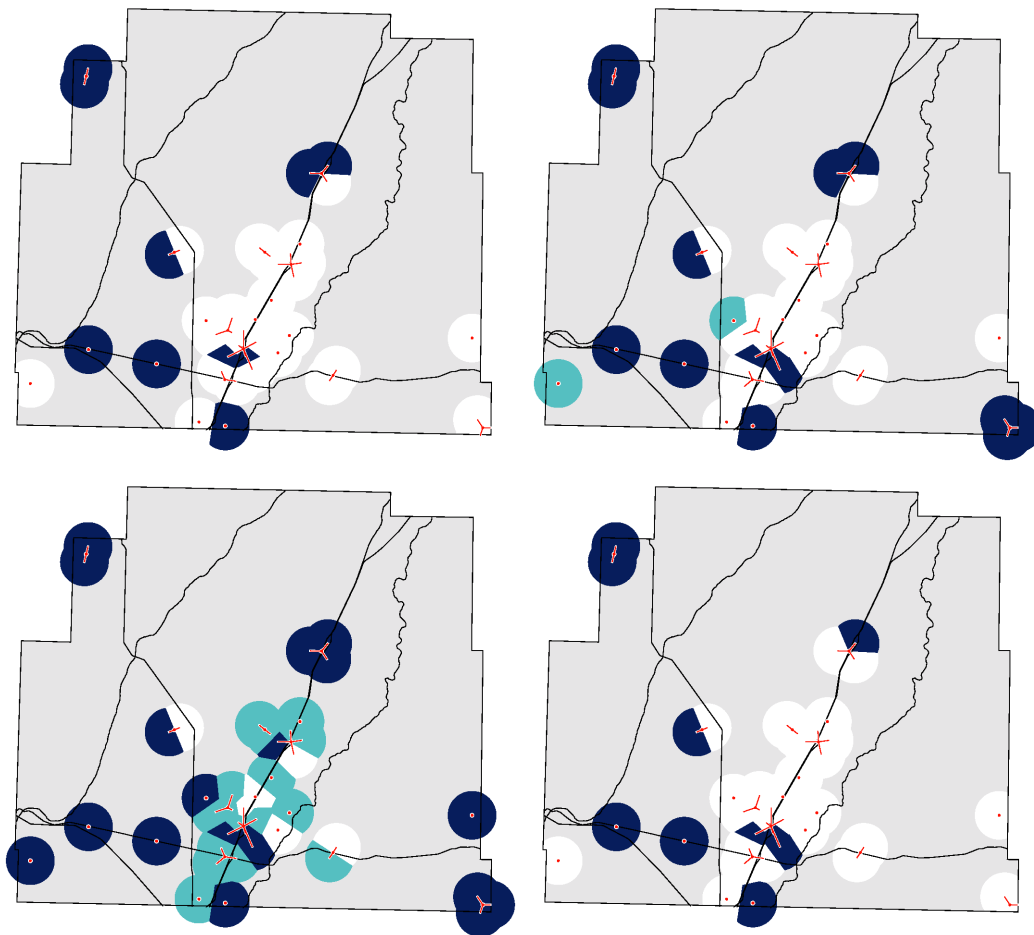
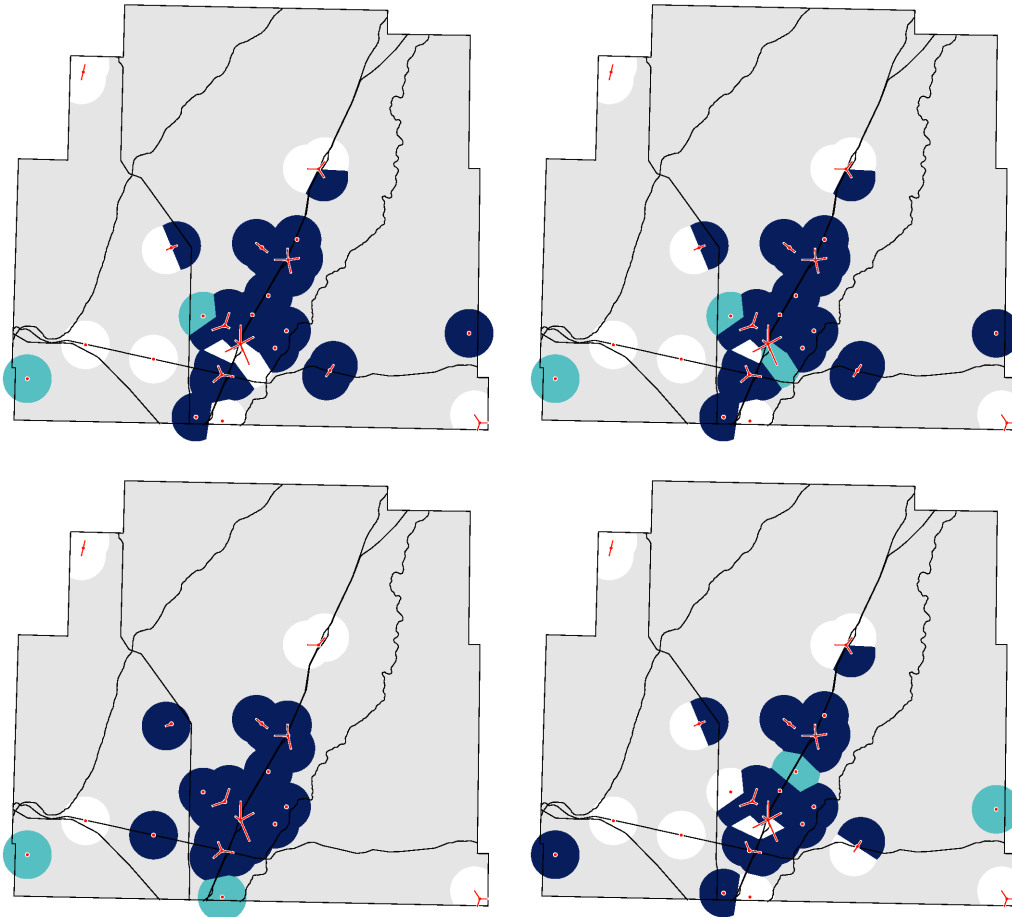


Figure 23. Distribution of phonemic-phonological cluster determinants, light green

cluster. Clockwise from top left:  $v_{\text{OEA}}$  ([ $\text{o}\text{ə}$ ] in blue),  $v_{\text{REALIZATIONUAPREVELAR}}$  (back realizations in blue, front realizations elsewhere),  $v_{\text{UEA}}$  ([ $\text{u}\text{ə}$ ] in blue),  $v_{\text{REALIZATIONFINALREN}}$  (non-metathesized *-re(n)* in blue, metathesized *-er(n)* elsewhere).



*Figure 24.* Distribution of phonemic-phonological cluster determinants, dark green cluster. Clockwise from top left:  $v_{\text{REALIZATIONUAPREVELAR}}$  (front realizations in blue, back realizations elsewhere),  $v_{\text{REALIZATIONOAPREVELAR}}$  (front realizations in blue, back realizations elsewhere),  $v_{\text{OEA}}$  ([ $\text{e}\text{ɔ}$ ] in blue),  $v_{\text{COULD2SRREDUCTION}}$  (reduced *ku's(t)* in blue, non-reduced *kunst* elsewhere).

Several determinants of these two clusters overlap with one another, but differ significantly in their associated features. Both the light and dark green clusters have  $v_{\text{OEA}}$  and  $v_{\text{REALIZATIONUAPREVELAR}}$  as determinants. As the above plots show, however, both clusters have different typical patterns of variant selection with respect to these two variables: where the light green cluster has [ $\text{o}\text{ə}$ ] for the former variable, the dark green cluster has [ $\text{e}\text{ɔ}$ ] (i.e., a fronted,

unrounded realization). Similarly, /ua/ diphthongs are generally realized before velar consonants with back rounded vowels among speakers in the light green cluster (e.g., *Buak* [buək] ‘book’, W52), where these same phonemes are realized with front rounded vowels among speakers in the dark green cluster (e.g., *Büak* [byək~bʏək] ‘book’). This pattern of pre-velar fronting extends even further in the dark green cluster, with /oa/ diphthongs also appearing in fronted form before velars in  $v_{\text{REALIZATIONOA}}v_{\text{PREVELAR}}$  (e.g., *Wöage(n)* [veo̞y̞ə(n)~vøo̞y̞ə(n)] ‘wagon’, W39). By contrast, speakers in the light green cluster do not generally have front realizations of these vowels, or indeed, any phonemes in the front rounded vowel space. Rather, the absence of [y~ʏ], either in pre-velar environments or in /u/ or /ua/ phonemes, appears to be a distinguishing feature of this cluster.

Taken together, these distinctive features give a clearer sense of the phonemic and phonological characteristics of each group:

- Catholic speakers in the dark blue cluster are immediately distinguishable by their monophthongal realizations of *ee* as [eː] and *oo* as [oː], as well as their realization of *ea* as [eə]. None of these variants are attested among any of the Mennonite contributors. At the same time, the Catholic group shares several other phonological features with the peripherally located Mennonite speakers, including the realization of /u/ as [u] (rather than [y]) and the absence of pre-velar fronting ( $v_{\text{REALIZATIONUA}}v_{\text{PREVELAR}}$ ,  $v_{\text{REALIZATIONOA}}v_{\text{PREVELAR}}$ ) and any metathesis in final *-re(n)* ( $v_{\text{REALIZATIONFINALREN}}$ );
- Peripherally located Mennonite contributors in the light green cluster appear to form one end of a spectrum of Mennonite Plautdietsch varieties, sharing few of their distinctive phonological variants with members of the central cluster. Thus, as noted above, speakers in this cluster show little metathesis of final *-re(n)*, and have back vowel realizations of /u/ and pre-velar *oa* and *ua*—all features that differ considerably from speakers in the central cluster;
- By contrast, centrally located Mennonite speakers in the dark green cluster represent the other extreme of Mennonite Plautdietsch varieties in the Saskatchewan Valley. Among speakers in this cluster, fronted realizations of /u/ and *oa* and *ua* in pre-velar environments are universal, reduction in certain modals (e.g.,  $v_{\text{COULD2SRREDUCTION}}$ ,  $v_{\text{CAN2SRREDUCTION}}$ ) is common, and *eiw* in  $v_{\text{AUUWEIW}}$  prevails;

- Mennonite speakers in the small, ‘transitional’ cluster in light blue share most major phonological features with the larger, centrally located group of speaker, often showing pre-velar rounding and a fronted realization of /u/. Where these speakers more often differ from the central cluster is in the *variability* of their selection of these features. Whereas speakers in the central cluster show essentially absolute uniformity in pre-velar fronting, reduction in certain modals, and fronted realizations of /u/, speakers in this transitional cluster vary between these realizations and those found in the peripheral group.

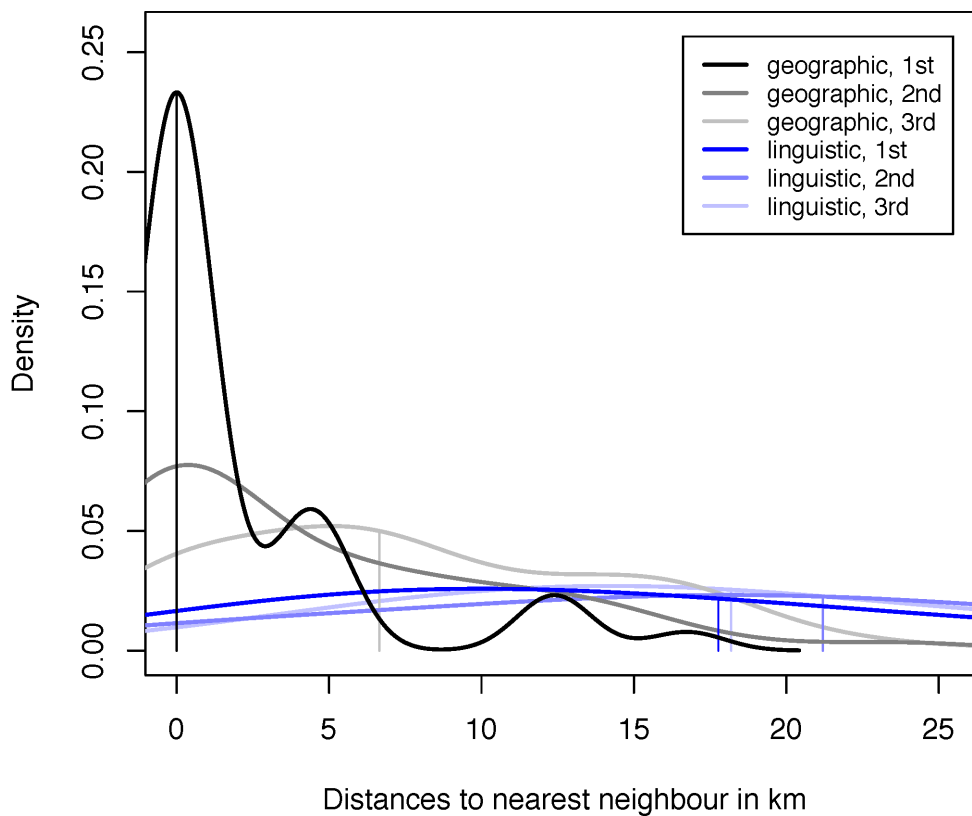
In sum, there is little evidence of an immediate correlation between phonological similarity and geographical distance in the Saskatchewan Valley—speakers do not generally share the greatest number of phonological features with their nearest geographical neighbours, as in a traditional dialect continuum. Nevertheless, there is evidence that Mennonite speakers with broadly similar phonological conventions predominate in certain areas, albeit with significant internal variation. These phonological patterns divide the local Mennonite community most immediately into central and peripheral groups, with a smaller group of ‘transitional’ Mennonite speakers sharing phonological features of both groups (although showing closer overall alignment to the features of the central cluster, despite their more disparate geographical distribution). Several phonological features also distinguish these Mennonite Plautdietsch speakers from their Catholic counterparts, who demonstrate marked linguistic differences in several respects. Section 5.3 will return to reconsider these divisions in light of the demography and settlement patterns of the Saskatchewan Valley.

### 5.2.2 *Lexical clustering*

The phonemic and phonological variables in the previous section provide initial evidence for several clusters of contributors who share similar patterns of variant usage, and suggest notable differences between both geographical areas of the Saskatchewan Valley (central vs. peripheral) and between Mennonite and Catholic Plautdietsch speakers. Since these classifications are, at best, working hypotheses based on observations in a single aspect of linguistic organization, it is useful to bring variation in areas other than the patterning of sounds into consideration, as well, to assist in assessing the overall plausibility of these conclusions.

This section does so by concentrating on variation related to lexical features, drawing on the 55 lexical variables coded in the *Fibel* Corpus, of which 2,318 instances are available.

It was noted earlier that geographical and linguistic distance did not correlate in a straightforward way for phonemic and phonological variables. Even so, it may be possible that lexical variation is different in this respect, and has a more predictable relationship with geographical distance. This hypothesis can be evaluated by considering in Figure 25, a density plot of the physical distances between nearest geographical and linguistic neighbours for all of the available lexical variables in the corpus.



*Figure 25.* Density plot of distances to nearest geographical and linguistic neighbours, lexical variables.

As with the phonemic and phonological variables of the preceding section, Figure 25 reveals little evidence of a general correlation between geographical proximity and linguistic similarity for lexical variables. While most contributors share their place of birth with at least one other

speaker (as is evident from the map in Figure 6), this does not appear to result in similar patterns of lexical variant selection. Rather, as this density plot makes clear, the distribution of geographical distances between pairs of speakers who are most alike in their lexical preferences is essentially flat, suggesting that there is little overall relationship between geographical distance and lexical similarity: some speakers' nearest lexical neighbours are close at hand, while others' are a considerable distance away, with no marked tendency towards either option.

Even if a general correlation between geographical distance and lexical variant selection in the *Fibel* Corpus data is not immediately apparent, there may be more coherent patterns to be found in contributors' lexical preferences. To consider whether or not lexical variation between speakers in this sample may contain further structure worth investigating, we begin with an MDS analysis of the feature distances between speakers for all lexical variables, as summarized in the two-dimensional plot in Figure 26.

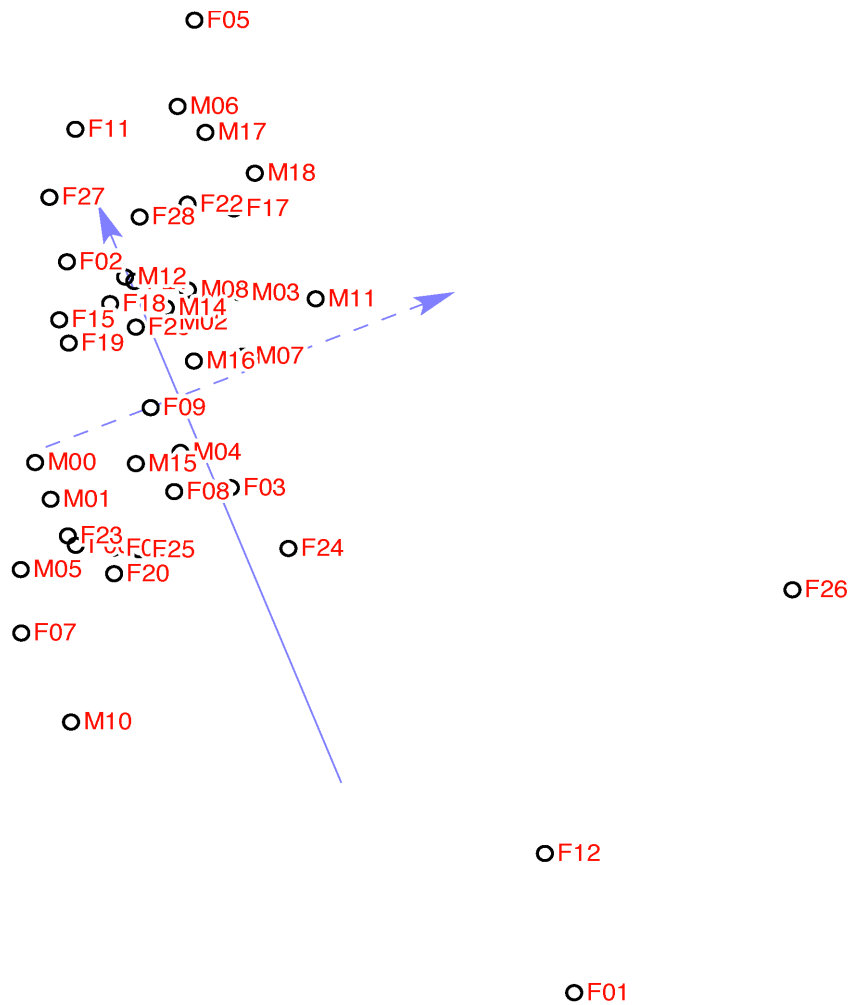


Figure 26. Two-dimensional MDS plot, lexical variables ( $r = 0.91$ ).

As with the MDS analysis of phonemic and phonological variation in Figure 15, two features of lexical variation among these speakers are immediately apparent from this plot. First, there is a notable difference between the three Catholic Plautdietsch contributors (F01, F12, F26) and the remaining Mennonite Plautdietsch speakers in their lexical preferences, as is evident from the distance between both groups in this plot. Second, among the Mennonite Plautdietsch speakers, there again appears to be some gradation of variation, represented here more or less coherently along the vertical axis. No deep lexical divisions are apparent at this level of abstraction among the Mennonite speakers in the corpus. Rather, while considerable differences exist between certain pairs of speakers (e.g., M10 and F05, who appear at opposite ends of the vertical axis), in

general, such differences appear to exist in a continuum, with few sharp breaks that would suggest more radical linguistic divisions between constituent groups of Mennonite Plautdietsch speakers.

This continuous distribution of lexical variation across the Mennonite contributors in the *Fibel* Corpus can be seen more clearly in Figure 27, which compares the MDS analysis given above with the corresponding hierarchical clustering of speakers by their lexical features.

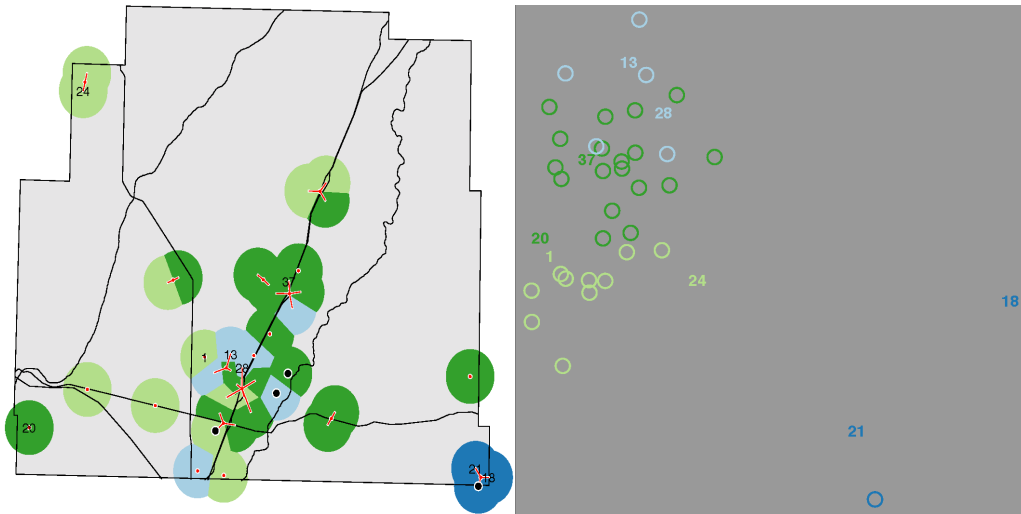


Figure 27. Hierarchical clustering by lexical features (Ward's method, four clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.91$ ; right).

While the cluster of Catholic speakers identified in the phonemic-phonological analysis and in the preceding MDS plot for lexical variables remains stable, appearing here in dark blue, the ‘transitional’ cluster noted in Section 5.2.1 is not evident in this analysis of lexical variation. Instead, these speakers form part of a larger cluster in the central Saskatchewan Valley, represented here in dark green. Unlike in the phonemic-phonological analysis, however, this central group is not monolithic: there is some evidence for additional, distinctive variation in this area that emerges as the cluster of speakers shown in light blue. This group occupies a geographical area similar to the larger cluster in dark green, and shows little definite separation from other clusters in the corresponding MDS analysis. These observations suggest that this division is more tenuous than the others thus identified, and that it may be reasonable to treat this group as forming part of a single macro-cluster in the central Saskatchewan Valley, with



members of the light blue cluster being the most lexically divergent from other, non-central Plautdietsch speakers in the sample.

While the appropriate level of granularity in this aspect of lexical clustering remains open to question, it is clear from the dendrogram in Figure 28 that the peripherally located group of Mennonite speakers differs substantially from other, centrally located Mennonites in their lexical preferences. On the whole, the peripheral group of speakers is much closer to the Humboldt Catholic group, but with considerable lexical differences still separating these two clusters.

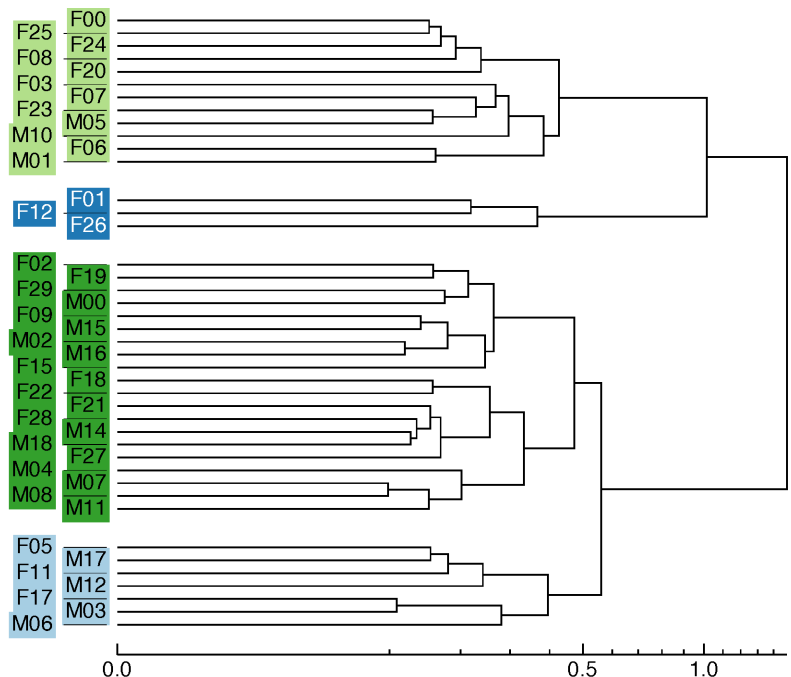


Figure 28. Dendrogram of hierarchical clustering by lexical features (Ward's method, four clusters).

While this high-level clustering suggests that coherent lexical differences exist between Saskatchewan Plautdietsch speakers, it would be helpful to know which particular linguistic features distinguish these groups from one another. Drawing on the distance-based measures of cluster determinants seen in Section 5.2.1, it is possible to identify the most distinctive features of each group of speakers, along with measures of the degree to which each feature is shared by all members of this group (within distance) and is not found among the members of other groups (between distance). Table 20 presents the highest-ranked lexical cluster determinants for each group of Plautdietsch speakers, as well as the within-distance and between-distance scores for

each determinant.

<i>Dark blue</i>	<i>Light green</i>	<i>Dark green</i>	<i>Light blue</i>
vHOUSE (3.41, -0.32, 3.08)	vUNTIL (1.43, -1.03, 0.40)	vTHATDAY (1.14, -0.87, 0.27)	vUNTIL (1.18, -1.03, 0.15)
vUNCLES (2.92, -0.40, 2.52)	vENGLISH (1.40, -1.09, 0.31)	vMOON (0.73, -0.59, 0.13)	vENGLISH (1.00, -0.78, 0.21)
vAUNTS (2.92, -0.40, 2.52)	vBECAUSE (1.10, -1.12, -0.02)	vSOME (0.68, -0.45, 0.22)	vGIRLS (0.96, -0.94, 0.02)
vLATE (2.89, -0.40, 2.49)	vGIRLS (0.79, -0.51, 0.28)	vWATERMELON (0.67, -0.59, 0.09)	vSAY (0.88, 0.20, 1.08)

*Table 20.* Distance-based cluster determinants, lexical variables.

The distance measures for each of these clusters provide useful information about the nature of the lexical differences that separate each group of speakers. For the Catholic Plautdietsch speakers in dark blue, cluster determinant scores are on the whole much higher and are accompanied by both low within distances (i.e., speakers in the cluster are generally quite similar to one another in their lexical preferences) and high between distances (i.e., speakers in the cluster have lexical preferences that are generally quite different from the speakers in other clusters). This supports the conclusion drawn from the MDS analysis above that the Catholic contributors are linguistically quite different from the non-Catholic contributors to the corpus, but also suggests a degree of internal coherence between these groups of speakers that was less evident in the above plot.

In the remaining clusters of Mennonite contributors, the between distances for each cluster determinant are generally low, especially when compared with the scores registered for the Catholic speakers in the dark blue cluster. This implies that many of the determinant lexical features of each cluster are not exclusive to members of that group, but rather are in use in other groups, as well. More commonly, what distinguishes one cluster from another is not the uniqueness of a given variant within that group (as would correspond to high between-distance scores), but rather the degree to which speakers in that group share similar patterns of variant usage (as indicated by the low within-distance scores). As we will see, this general pattern of variants being shared across many clusters, but having different conventional patterns of usage in each one, recurs in other aspects of linguistic variation in the Saskatchewan Valley, as well, and will be revisited in more detail later on.

As before, distribution plots of cluster determinants help elucidate the structure of lexical differences among these groups of speakers. Beginning with the cluster in dark blue, Figure 29 presents the geographical distribution of the highest-ranked lexical cluster determinants for the Catholic Plautdietsch speakers. Inspection of these plots reinforces the impression of marked linguistic differences between these speakers and others in the Saskatchewan Valley: all of the lexical variants associated with these speakers for the highest-ranked cluster determinants are present among members of this group, and none of these variants are attested among speakers in other groups.

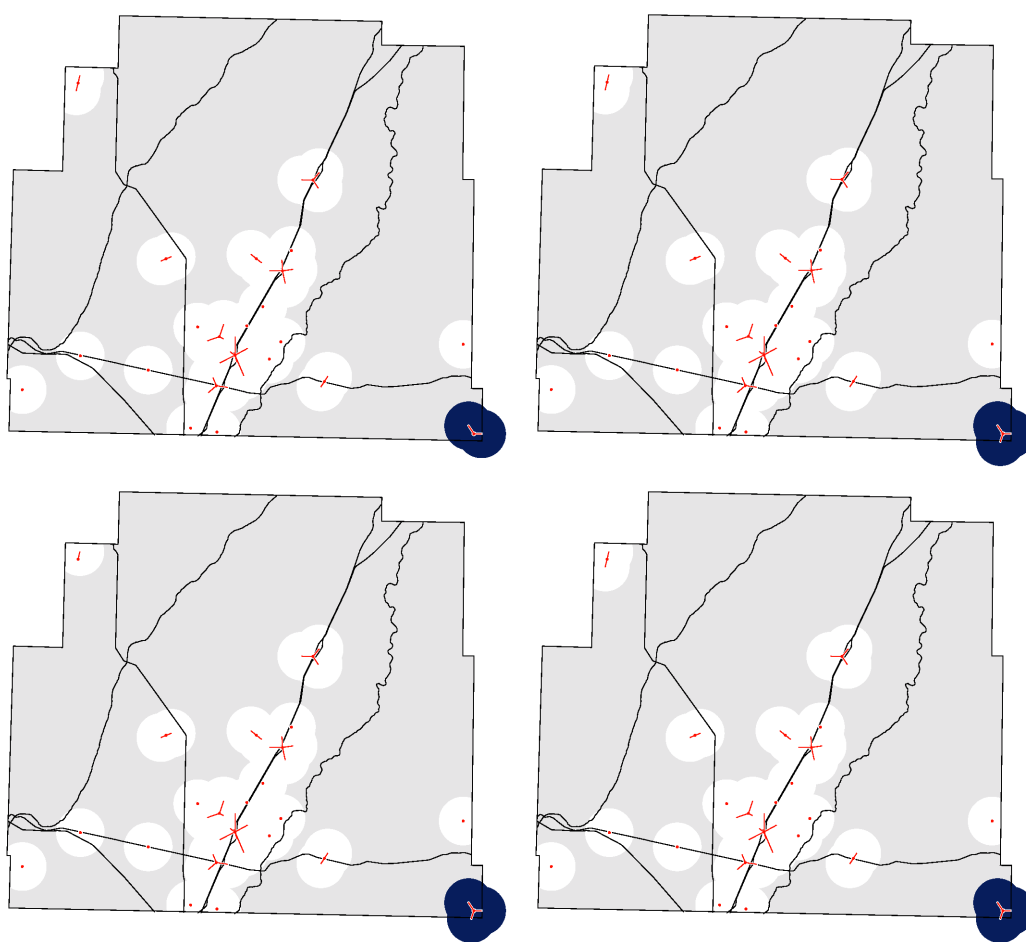


Figure 29. Distribution of lexical cluster determinants, dark blue cluster. Clockwise from top left:  $v_{\text{HOUSE}}$  (*Kot* in blue, *Huus* elsewhere),  $v_{\text{UNCLES}}$  (*Oohms* in blue, *Onkels* elsewhere),  $v_{\text{AUNTS}}$  (*Mumms* in blue, *Tauntes* elsewhere),  $v_{\text{LATE}}$  (*spod* in blue, *lot* elsewhere).

Indeed, several of these items are not attested with these meanings in Mennonite varieties of Plautdietsch, if they are attested at all. For example, *spod* ‘late’ is not reported in either Rempel (1995) or Thiessen (2003) (although the latter does have *spoda*, *spohda* ‘later’), while both lexicographers define *Kot* not as a house, as the Catholic Plautdietsch speakers consistently report, but more specifically as a ‘small, poor farm cottage’ (Thiessen) or a ‘small, simple abode’ (Rempel). For both *Oohm* ‘uncle’ and *Mumm* ‘aunt’, the situation is somewhat more complex. While Rempel includes ‘uncle’ in his definition for *Oom*, he also includes ‘sir, Mr., reverend’ as equivalent translations. Thiessen does likewise, defining the term as ‘Mister, Reverend, sir’. Unlike the Catholic contributors, who reported that this was their only term for ‘uncle’ (and who considered *Onkel* to be a borrowing from High German), many Mennonite speakers appear to have additional associations with *Oohm* that are not restricted to kinship relations. As M18 comments, the term *Oohm* was reserved in some families to refer specifically to ‘great uncles’ or to pastors:

- (11) *See*, (...) wan ‘et ‘en, ‘en *great uncle* wea, // dan säd’ wi “Oohm.” (...) Ooda wan daut ‘en, ‘en Prädja wea, dan send daut uk “Oohms.”  
 ‘See, if it were a great uncle, then we said *Oohm*. Or if it were a pastor, then those are *Oohms*, too.’  
 (M18, 2011-08-02, 21m09s665–21m25s480)

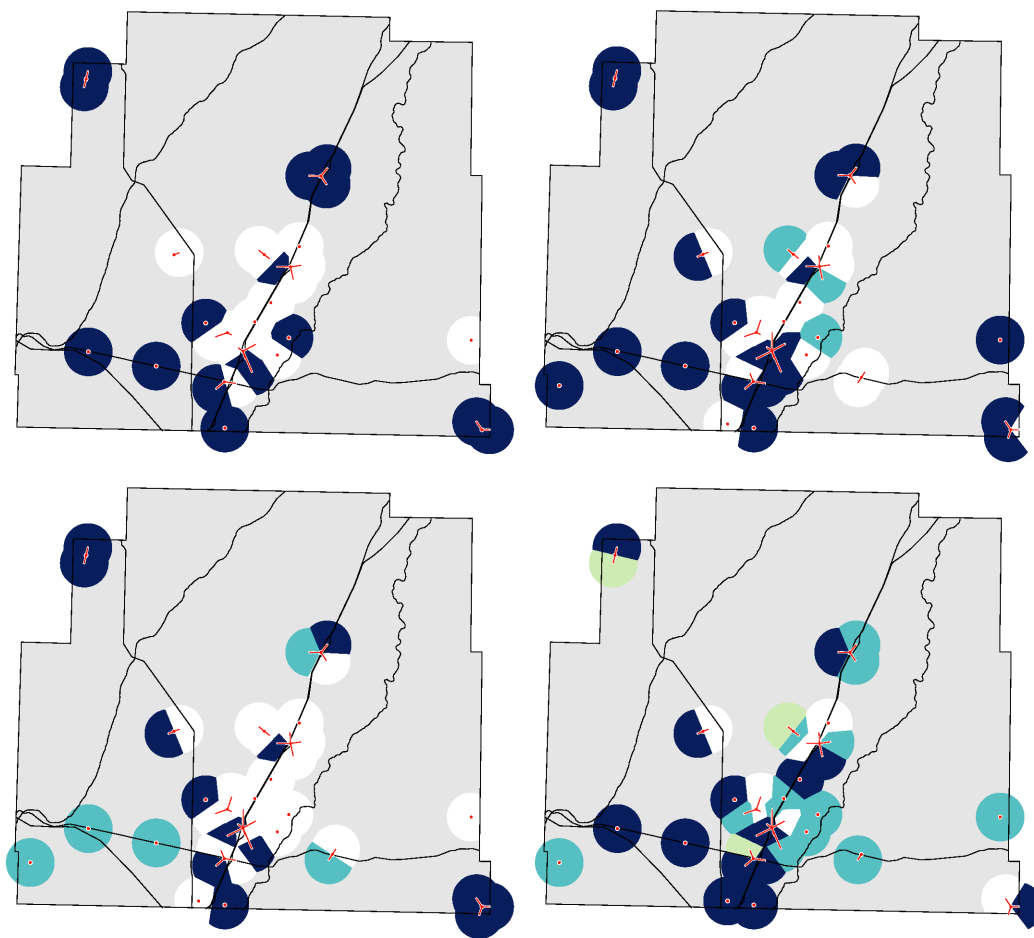
This is confirmed separately by M06, who provides an example of how the terms *Oohm* and *Mumm* were used in his family:

- (12) M06: My great uncle, we called him *Oohm Peeta*.  
 CDC: *Oohm Peeta*.  
 M06: Uh-huh.  
 F14: And how, what was the aunt?  
 M06: Oh... // but that was, like, *Auntje-mumm*.  
 (2011-08-07 (01), 17m28s950–17m39s580)

While both Rempel and Thiessen define *Mumm* as ‘aunt’, Thiessen further notes that this usage is restricted to Mexican Mennonites and Old Colony Mennonites, suggesting a more limited use of this term than would be expected from its English translation. Along with M06 in the example in (12), several contributors to the *Fibel* Corpus commented that the term *Mumm* was used specifically in terms of address (e.g., *Auntje-mumm* ‘Aunt Anna’), often for great-aunts,

rather than as a free-standing word in its own right, as in the dark blue cluster above.

Examples such as these suggest a more substantial lexical division between the Mennonite and Catholic contributors to the *Fibel* than was initially apparent—yet one that dialectometric measures such as those applied above identified on the basis of their distinctiveness in the dark blue cluster. By comparison, the lexical cluster determinants among the speakers in the remaining groups show less abrupt differences, though closer consideration of these features also proves revealing. Figure 30 shows the geographical distribution of the four highest-ranked determinants of the light green cluster.



*Figure 30.* Distribution of lexical cluster determinants, light green cluster. Clockwise from top left: vUNTIL (*bat* in blue, *bott* elsewhere), vENGLISH (*Englisch* in blue, *Engelsch* elsewhere), vBECAUSE (*wiels* in blue), vGIRLS (*Mäakjes* in blue, *Me(r)jalles* and *Me(r)jalle(n)* elsewhere).

Although the differences between this cluster and its neighbours are less striking, several features of this plot bear noting. First, as in the dendrogram in Figure 28, the speakers in this cluster generally appear to be most similar to the members of the Humboldt Catholic cluster in the bottom right-hand corner of the plots. In most aspects of their variant selection for these highest-ranked variables, speakers in the light green cluster have much the same preferences as members of the Catholic cluster. This suggests that these Mennonite speakers more than others in the Saskatchewan Valley share certain lexical features with members of the Catholic group. With regard to the variants themselves, while all appear to differentiate this group from other Mennonite clusters, only the variable *vGIRLS* was frequently the subject of contributors' commentary. Unlike most variables in the present sample, this lexical feature appears to be socially marked. In (13), for instance, contributor M00 provided two translations of 'girls' in S01 ("The little girls shouldn't jump on the bed"), once as *Mäakjes* and once as *Mejalles*, which immediately prompted the following commentary:

- (13) M00: *Mejalles*, [...] *Mäakjes*.  
 CDC: Is there a difference there, or, uh...?  
 F20: *Na*, [...] "*Mejalles*" is a little more slang.  
 M00: Yeah, a little more crude.  
 F20: "*Mäakjes*" is a little more proper.  
 (2011-08-09 (02), 00m29s906–00m40s570)

For at least some contributors to the corpus, *Mejalles* appears to have mildly negative associations, as is reflected in Thiessen (2003, s.v. *Mejahl*, *Merjahl*), who comments that this term is "slightly pejorative in some circles", but that "in many Mennonite villages the term *Mejahl* is the only term used for a girl". For other contributors, however, no such negative connotation is apparent. For M14, while *Mäakjes* is a familiar term, his preference is decidedly *Mejalles*:

- (14) "*Mäakjes*" es, uh, // woat föaken jebrukt en'e Städ "*Mejalles*." // Ekj [...] wudd äwajens, // ekj wudd "*Mejalles*" saijen.  
 '*Mäakjes* is, uh, // is often used instead of *Mejalles*. // Anyway, I would, // I would say *Mejalles*.'  
 (M14, 2011-08-03 (01), 00m22s835–00m31s565)

Interestingly, exactly the opposite selection of variants for several of these features (i.e., *vUNTIL*,

vENGLISH, vGIRLS) characterizes the light blue cluster in Figure 31. While some of these variants are also shared with members of the closely related dark green cluster, the stark contrast between this cluster and the one above is noteworthy, suggesting a greater degree of lexical divergence between these two groups than between other subsets of Mennonite speakers.

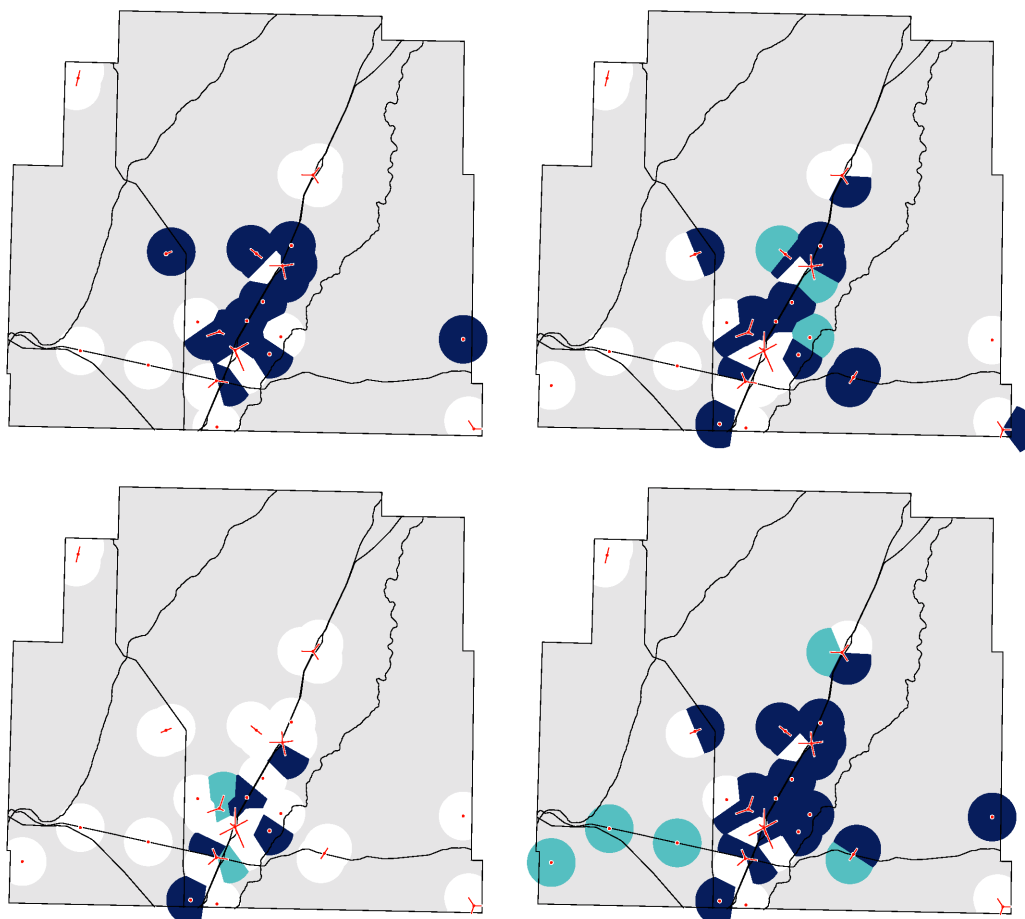


Figure 31. Distribution of lexical cluster determinants, light blue cluster. Clockwise from top left: vUNTIL (*bott* in blue, *bat* elsewhere), vENGLISH (*Engelsch* in blue, *Englisch* elsewhere), vGIRLS (*Me(r)jalle(n)/Me(r)jalles* in blue, *Mäkjes* elsewhere), vSAY (*saje(n)* in blue, *saije(n)* elsewhere).

Finally, the distribution of the highest-ranked determinants of the largest cluster, shown earlier in dark green, are displayed in Figure 32 below. As noted above and suggested by the previous dendrogram, this cluster demonstrates the greatest lexical similarity with the light blue cluster, sharing many of their characteristic variants in common. These two groups pattern together

lexically, with members of the light blue cluster having fewer features in common with speakers in the light green and dark blue clusters, and members of this dark green cluster having slightly more. In both cases, the primary division remains between the peripherally located Mennonite and Catholic Plautdietsch speakers and these groups in the centre of the Saskatchewan Valley, although this analysis suggests that a range of subtler differences may exist within each of these larger clusters.

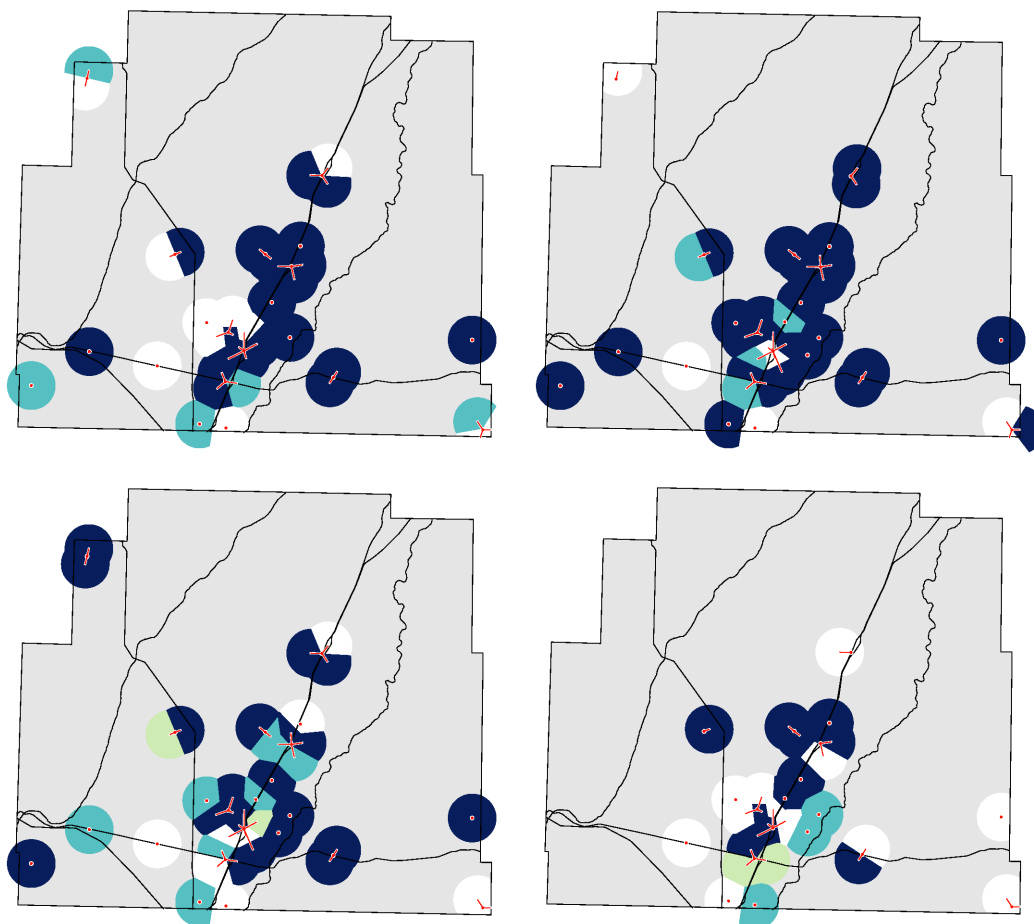


Figure 32. Distribution of lexical cluster determinants, dark green cluster. Clockwise from top left:  $v_{\text{THATDAY}}$  (*dän Dag* in blue, *däm/dee Dag* elsewhere),  $v_{\text{MOON}}$  (*Mon* in blue, *Mond* elsewhere),  $v_{\text{SOME}}$  (*waut* in blue),  $v_{\text{WATERMELON}}$  (*Rebuus* in blue).

In sum, clustering on lexical features such as these contributes valuable information to the picture of variation begun in the preceding section in several respects:

- Lexical clustering suggests a four-way division between contributors to the *Fibel*



Corpus, with a marked, high-level split between Catholic and Mennonite Plautdietsch speakers, followed by less precipitous divisions between peripherally and centrally located Mennonite speakers in the Saskatchewan Valley. Among the Mennonite speakers, the sharing of cluster determinants across groups suggests much less dramatic clefts between groups of Mennonite speakers than between Mennonites and Catholics, with the peripherally located Mennonite speakers generally having more variants in common with Catholic speakers;

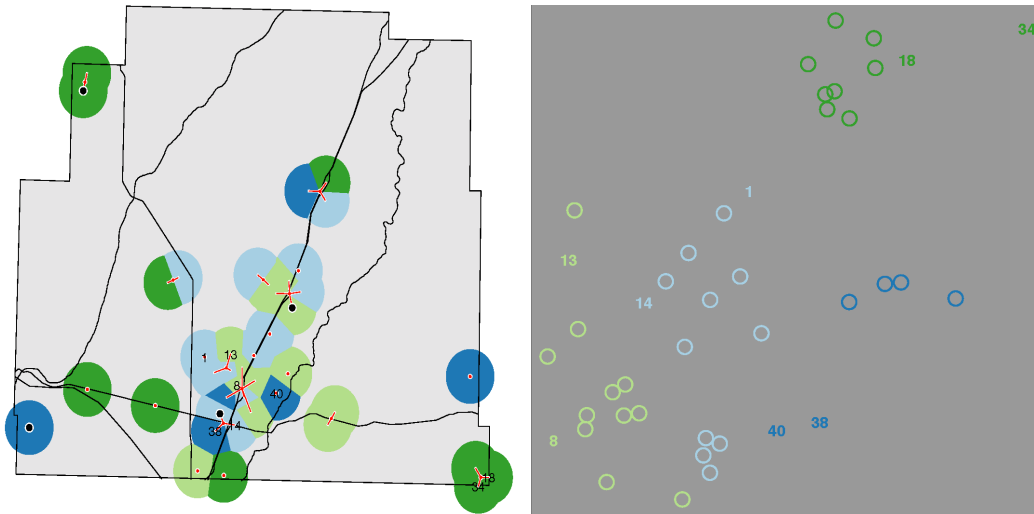
- The earlier ‘transitional’ group of speakers identified in the phonemic-phonological analysis is not apparent in this lexical analysis. Instead, those speakers pattern together here with larger groups of centrally located Mennonite speakers;
- In further contrast to the phonemic-phonological analysis, centrally located Mennonite speakers are divided here into two groups who share greater or lesser numbers of their lexical variants with peripheral Mennonite speakers. This division, while tentative, is not evident in phonemic-phonological clustering.

In general, while the above clustering methods successfully identify several distinct groups of Mennonite speakers, lexical variation in these clusters does not tend to be characterized by exclusive distributions of variants. Even with this being the case, the above methods suggest that coherent patterns of lexical difference between groups of Mennonite speakers are nevertheless present and discernible. This patterning of non-distinctive variants in distinctive constellations is noteworthy, and will be revisited later on below.

### 5.2.3 *Morphological clustering*

The recurrence of clusters across the preceding two analyses—sometimes with more internal divisions, as with the emergence of two ‘central’ groups of speakers in the above lexical clustering analysis, and sometimes with less, as with the merger of the earlier-identified group of transitional speakers into these central clusters—provides further motivation for considering morphological variables in a similar way. Although a smaller set than either the phonemic-phonological or lexical variables, the morphological variables available in the *Fibel* Corpus comprise 27 features, represented in 1,282 instances in the corpus. Figure 33 presents an initial hierarchical clustering analysis of these features, with an accompanying two-dimensional MDS

plot to assess the reliability of the posited groupings.



*Figure 33.* Hierarchical clustering by morphological features (Ward's method, four clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.93$ ; right).

In the context of the two preceding clustering analyses, several features of these morphologically based clusters merit discussion. First, as the above MDS plot makes clear, the morphologically defined clusters in this analysis can be divided into two large groups: one at the geographical periphery of the Saskatchewan Valley, and another, less tightly cohering, in the centre of the region, comprising several smaller divisions. Notably, the Humboldt Catholic speakers are included among the former group, patterning together with their peripheral Mennonite counterparts in their morphological features. This similarity is all the more striking when one recalls that the Catholic Plautdietsch speakers have displayed lexical and phonological traits in the preceding sections that have been markedly different from the larger Mennonite Plautdietsch population (although often still displaying greater overall similarity in these features with the peripheral Mennonite groups than with others, as is evident from the dendrogram in Figure 34 below).

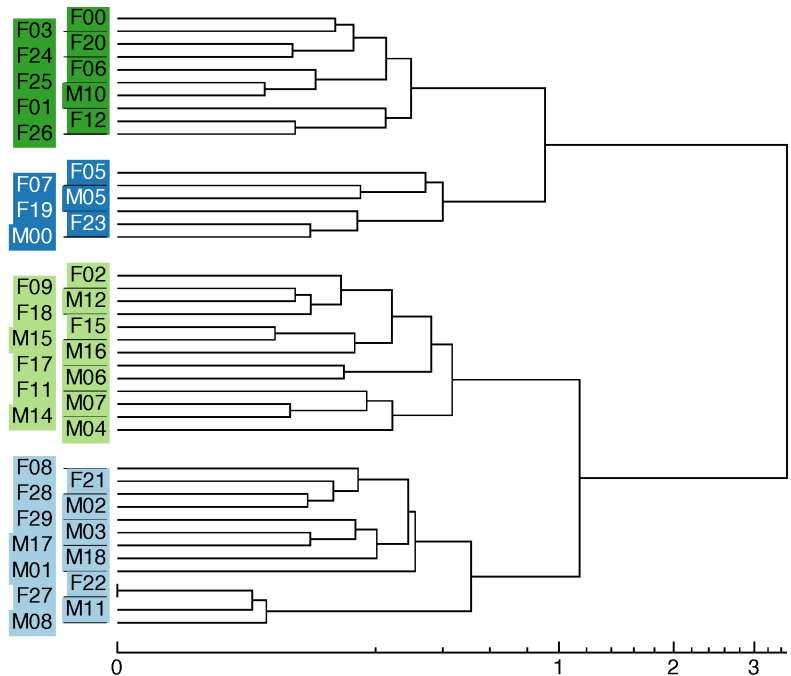


Figure 34. Dendrogram of hierarchical clustering by morphological features (Ward's method, four clusters)

Also notable is the recurrence of the transitional group of speakers who were first identified on the basis of phonemic-phonological features in Section 5.2.1. Even from the bird's-eye view afforded by this visualization, it is clear that these speakers maintain much the same east-west geographical distribution as was seen in the previous section, and that their morphological features bear some similarity to both the peripheral-Catholic and central groups (hence their positioning in the lower-right quadrant between both groups on the MDS plot). Although this clustering analysis suggests that the transitional speakers are most similar in their morphological features to the central group presented in light blue, there is little clear coherence in the central group that would support a geographical division between these speakers and the other centrally located cluster in light green, even if linguistic differences remain between them. In general, linguistic distance and geographical distance do not appear to be closely related for morphological variables, as the density plot in Figure 35 suggests.

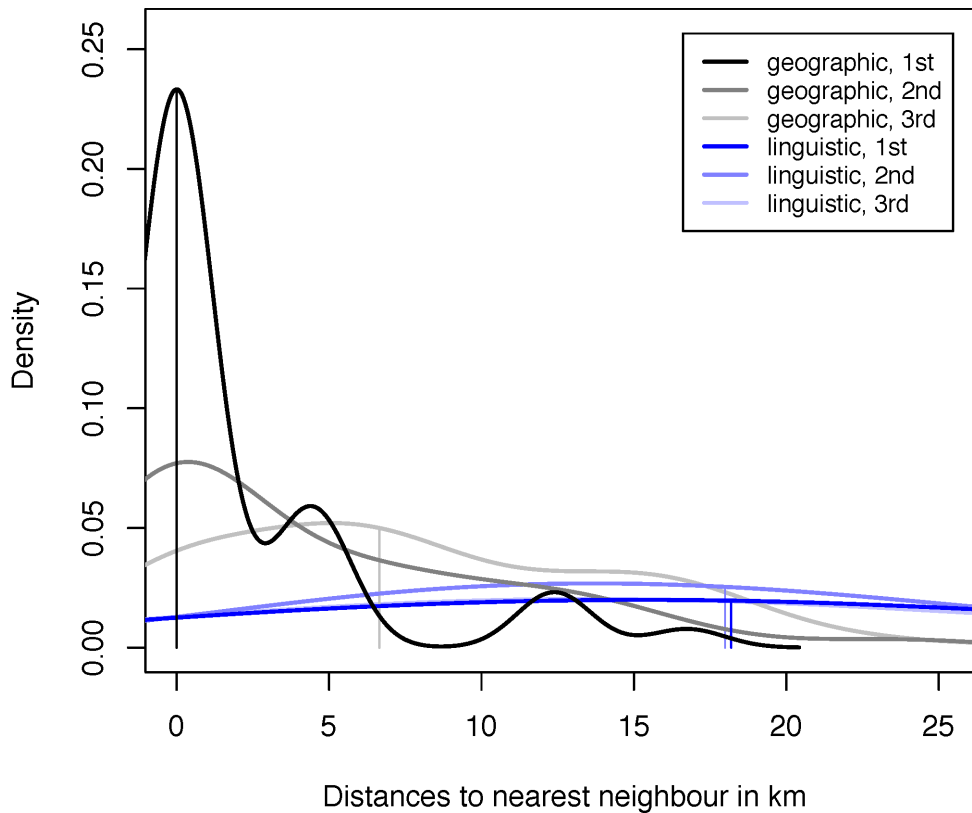


Figure 35. Density plot of distances to nearest geographical and linguistic neighbours, morphological variables.

Having a clearer impression of the distribution of morphological variation across contributors to the corpus, we may now turn our attention to understanding what particular features inform these divisions. Table 21 presents the highest-ranked cluster determinants for the available morphological variables in each of the four clusters.

<i>Dark green</i>	<i>Dark blue</i>	<i>Light green</i>	<i>Light blue</i>
vENNOUNSG (1.86, -1.03, 0.84)	vTHEACCMSG (0.98, -0.39, 0.58)	vENWITHOUT (2.04, -0.96, 1.08)	vEN3PL (0.85, -0.93, -0.09)
vEN3PL (1.86, -0.93, 0.93)	vENWITHOUT (0.82, -0.96, -0.14)	vENOTHER (2.04, -0.96, 1.08)	vENVERBAL (0.79, -0.90, -0.10)
vENPASTPART (1.82, -1.19, 0.63)	vENOTHER (0.82, -0.96, -0.14)	vENNOMINAL (1.33, -1.02, 0.30)	vENNOUNSG (0.75, -0.72, 0.03)
vENINF (1.46, -0.97, 0.49)	vENDINGADJACCMSG (0.73, -0.28, 0.46)	vENNOUNSG (1.14, -1.03, 0.11)	vENPASTPART (0.72, -0.62, 0.10)

Table 21. Distance-based cluster determinants, morphological variables.

From this table, it is evident that variables related to *-e(n)* predominate as morphological cluster determinants. With the exception of the transitional cluster in dark blue, where the accusative form of the masculine singular definite article ( $v_{\text{THEACCMSG}}$ ) and accusative masculine singular adjective endings ( $v_{\text{ENDINGADJACCMSG}}$ ) are among the highest-ranked morphological determinants, the most prominent linguistic dividing-lines between the remaining clusters lie in features related to variation in *-e(n)*. Even while the above summary statistics suggest a relatively complex distribution of these variables, inspection of the corresponding distribution plots makes their patterning more transparent. Beginning with the dark green cluster that encompasses both Catholic and peripherally located Mennonite contributors, Figure 36 shows the distribution of variants for the highest-ranked morphological determinants among these speakers.

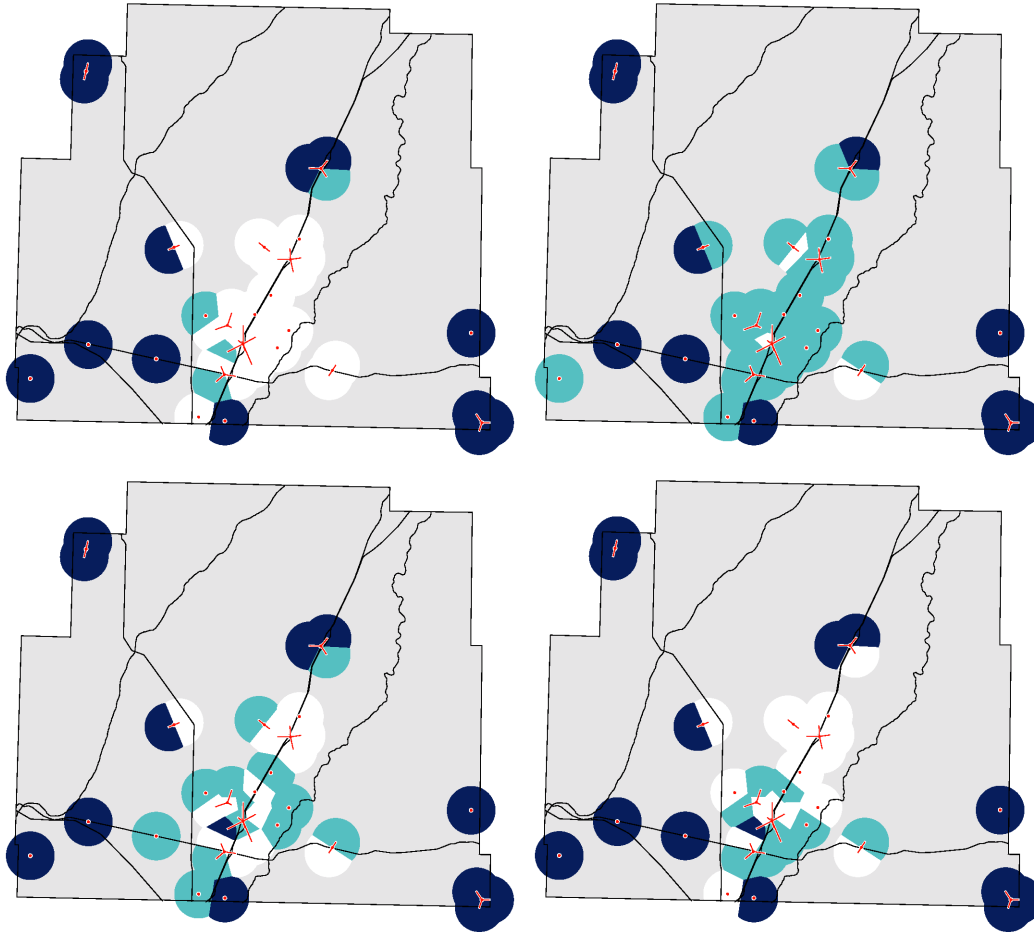


Figure 36. Distribution of morphological cluster determinants, dark green cluster. Clockwise from top left:  $v_{EN}nOUNSG$ ,  $v_{EN}3PL$ ,  $v_{EN}PASTPART$ ,  $v_{EN}INF$  ( $-e$  in blue).

These distribution plots make apparent what the Catholic and Mennonite members of this cluster have in common: both groups of speakers share the same, essentially categorical preference for  $-e$  in  $-e(n)$  variables. While members of the remaining clusters (with the exception of the transitional cluster; see below) vary between use of  $-e$  and  $-en$  variants (as indicated by the points in cyan in these distribution plots), no such variation is found among speakers in this dark blue cluster.

This categorical use of  $-e$  among Catholic and peripherally located Mennonite Plautdietsch speakers, as is central to the morphological profile of this cluster, stands in striking contrast to the apparent variability of these same features in the remaining clusters. The observation of intra-speaker variability in these features is significant in its own right: differences

in *-e* and *-en* have long been treated as one of the most important points of linguistic divergence between Mennonite Plautdietsch speech communities, both in the linguistic literature and among Plautdietsch speakers (Quiring 1928: 53, fn. 68; Dyck 1964: 66; Jedig 1966: 84; Tolksdorf 1985: 327; Epp 1987: 67; Moelleken 1987: 180–181; Epp 1993: 78; Nyman 1997: 267; Loewen 1998: 144; a.o.). With few exceptions, accounts of this variation have described this feature as basically categorical in its distribution across speakers: as with the contributors in the dark green cluster above, it is assumed that individual speakers employ either *-e* or *-en*, but not both. Yet, in the central and peripheral clusters, these features appear to be anything but categorically distributed: contrary to expectation, intra-speaker variation in *-e(n)* is clearly and repeatedly observable among these groups.

Delving deeper into examples of this variation in the *Fibel* Corpus suggests two possible causes for the observed intra-speaker variability, each associated with different clusters of speakers, and its importance for morphological clustering. Consider first the distribution plots in Figure 37, which present the distributions of the highest-ranked morphological cluster determinants for the transitional speakers represented in dark blue above.

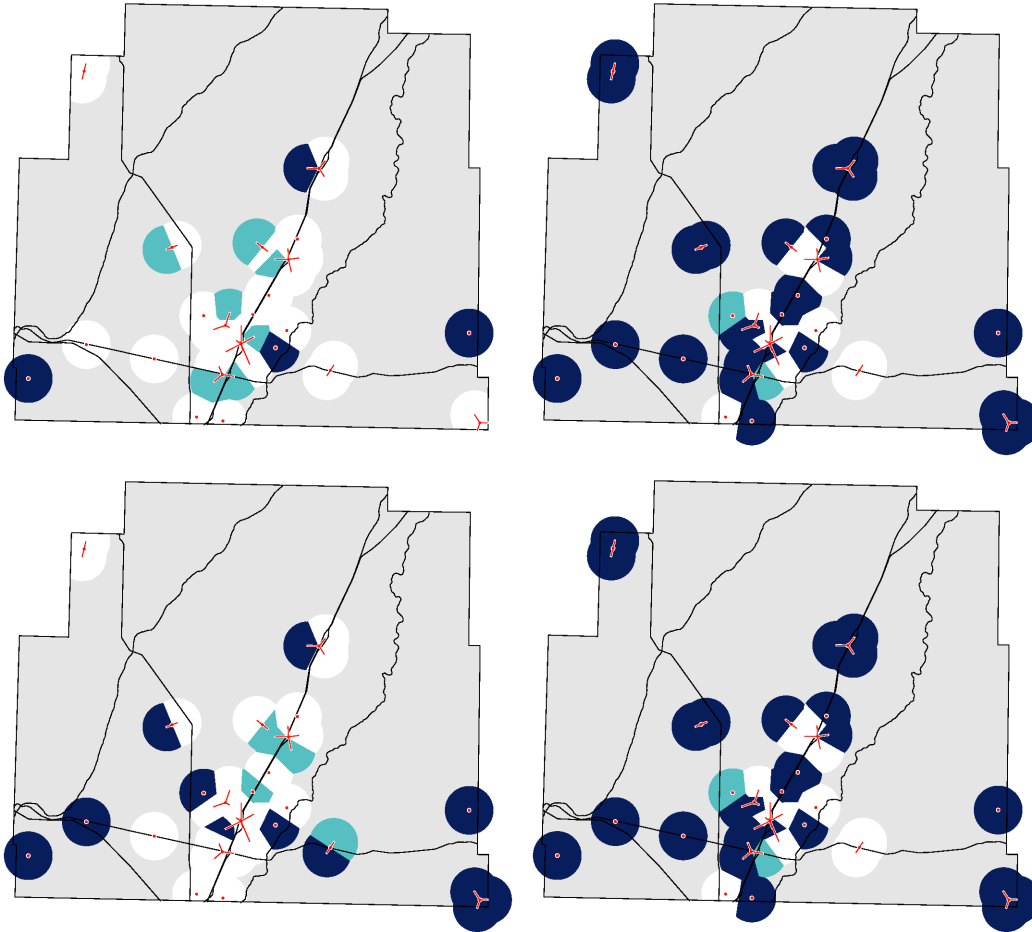


Figure 37. Distribution of morphological cluster determinants, dark blue cluster. Clockwise from top left:  $v_{\text{THEACCMSG}}$  (*dee* in blue, *däm/dän* elsewhere),  $v_{\text{ENWITHOUT}}$  (*-e* in blue),  $v_{\text{ENOTHER}}$  (*-e* in blue),  $v_{\text{ENDINGADJACCMSG}}$  (*-a* in blue, *-en* elsewhere).

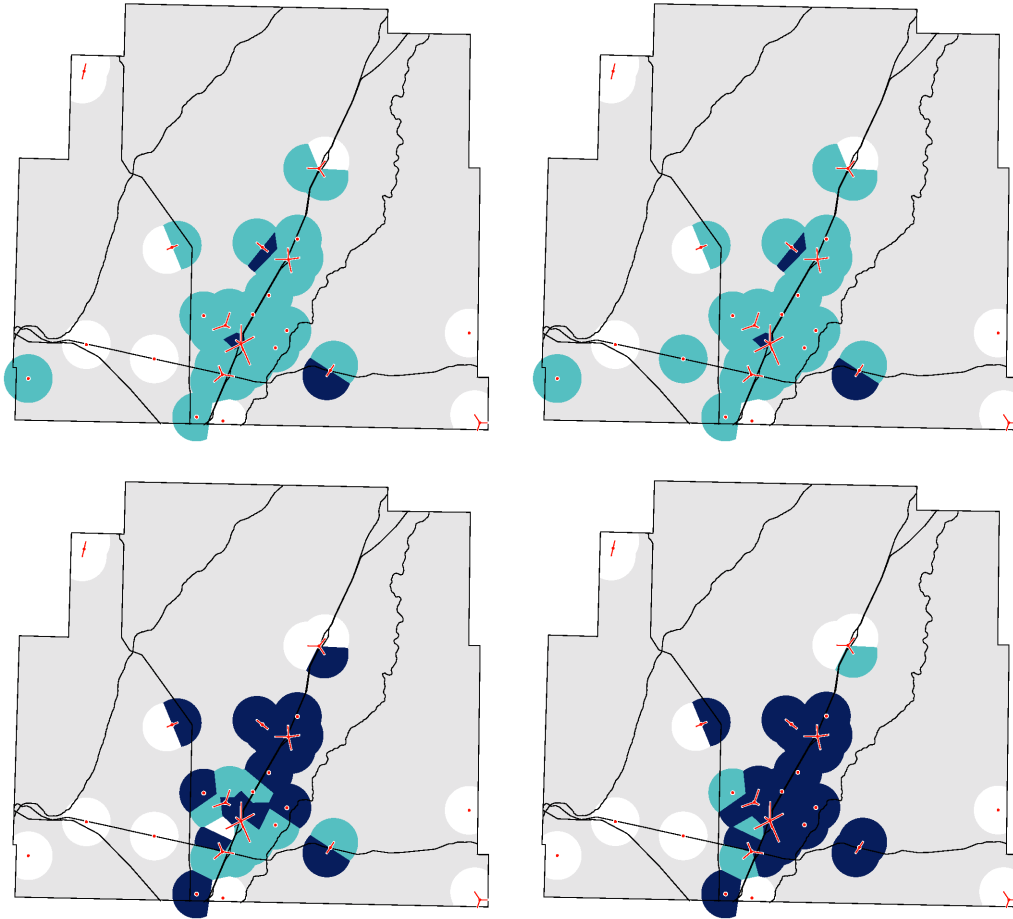
Although it is interesting that this transitional cluster should have non-*e(n)* variables among its top shibboleths, even these features have low between-distance scores that are apparent in the shared distributions seen in the above plots. This lack of distinctiveness from either the Catholic and peripheral Mennonite cluster or the remaining central clusters appears to be the defining feature of this group. Compared to speakers in other clusters, the morphological preferences of speakers in this group are not entirely black-and-white, as with dark green cluster of Mennonite and Catholic Plautdietsch speakers, nor as variable as the central Mennonite clusters, but rather show a distinctive mix of variant usage patterns, even as these variants are shared with other groups. This is evident in individual examples of *-e(n)* variation from members of this cluster, where both *-e* and *-en* may vary in the same utterance without apparent motivation. In the



translations of S05 (“Our boss says that we can have dinner early today”) given in (15), contributor M05 alternates between *äten* ‘to eat’ (with *-en*) in the first version, and *äte* (with *-e*) in the second, with both translations being otherwise identical:

- (15) a. Ons Bauss saigt, daut wi kjenne [.] vondöag [.] tiedig Meddag äten.  
 our boss says that we can.PL today early lunch eat.INF  
 ‘Our boss says that we can eat lunch early today.’  
 (M05, 2011-08-04 (02), 02m15s085–02m19s015)
- b. Ons Bauss saigt, daut wi kjenne vondöag // tiedig me-, Meddag äte.  
 our boss says that we can.PL today early HES lunch eat.INF  
 ‘Our boss says that we can eat lunch early today.’  
 (M05, 2011-08-04 (02), 02m19s045–02m25s055)

A different kind of variation is found among the members of the central cluster. Figure 38 and Figure 39 visualize the distribution of the highest-ranked morphological determinants of the light blue and light green clusters, respectively. All of these cluster determinants relate to *-e(n)* variation in different constructions. These plots make several features of this variation more readily apparent: first, and perhaps most immediately, variation in *-e(n)* endings is markedly different from construction to construction. While *-e(n)* variation related to nouns (vENNOUNSG) and past participles (vENPASTPART) in Figure 38 shows a much greater tendency towards exclusive use of *-en*, variation in third-person plural verb forms (vEN3PL) is decidedly more mixed.



*Figure 38.* Distribution of morphological cluster determinants, light blue cluster. Clockwise from top left:  $vEN3PL$ ,  $vENVERBAL$ ,  $vENNOUNSG$ ,  $vENPASTPART$  (-en in blue).

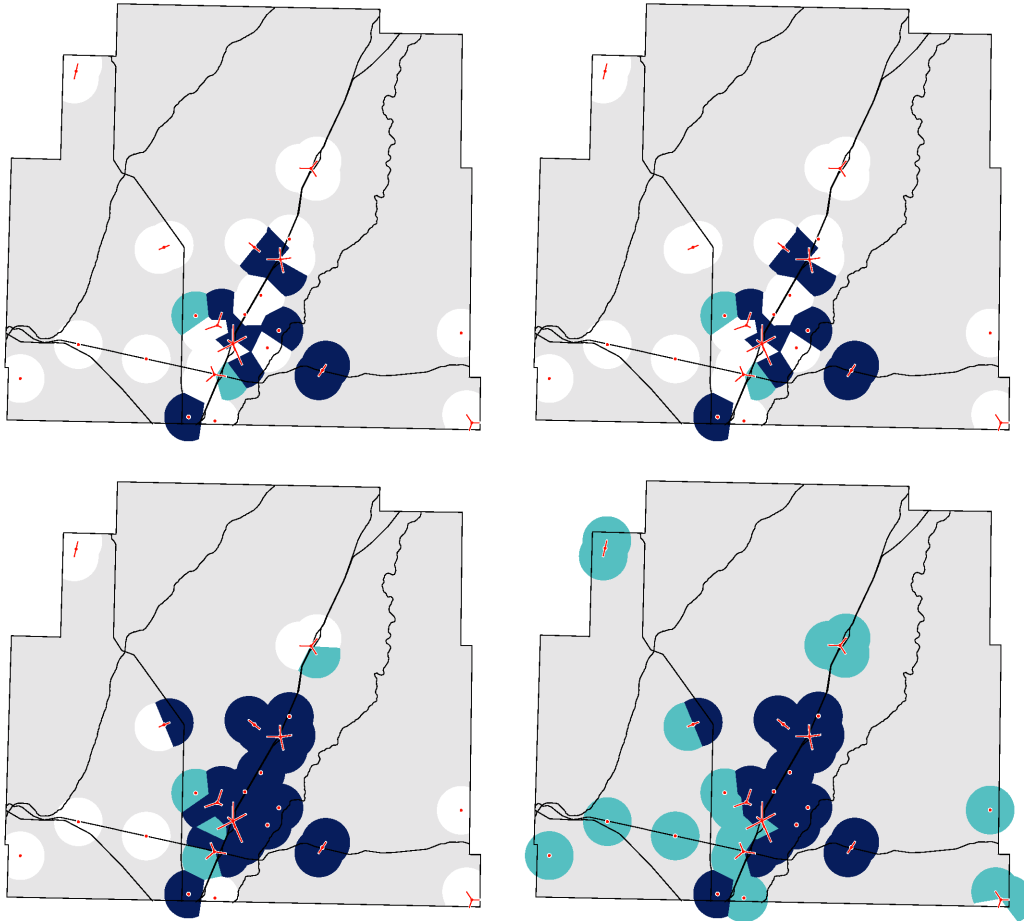


Figure 39. Distribution of morphological cluster determinants, light green cluster. Clockwise from top left:  $v_{EN}WITHOUT$ ,  $v_{EN}OTHER$ ,  $v_{EN}NOMINAL$ ,  $v_{EN}NOUNSG$  (-en in blue).

Also evident from these plots is the degree to which these features are shared across speakers in the central region of the Saskatchewan Valley. With the possible exception of  $v_{EN}PASTPART$ , where some differentiation between the light blue and light green clusters might be perceived, most central speakers appear to pattern in much the same way. Given apparent differences in  $-e(n)$  variation, both between constructions and between central and transitional speakers, it is worth considering instances of this variation in the central group in more detail. In (16), several renderings of S10 (“The cat sat beside the door that was getting painted”) provided by contributors in the central cluster demonstrate unexpected variation in  $-e(n)$  in the past participle *jeforwe(n)* ‘(have) painted’:

- (16) a. De Kaut, dee saut // besied de Däa, // dee sull // jeforw e woaren.  
 the cat it sat beside the door REL should painted get.INF  
 ‘The cat, it sat beside the door that was supposed to get painted.’  
 (F08, 2011-08-08 (01), 05m26s725–05m36s655)
- b. De Kaut saut bisied de Däa, waut doa jeforw e word.  
 the cat sat beside the door REL there painted got  
 ‘The cat sat beside the door that got painted.’  
 (F09, 2011-08-03 (01), 04m31s175–04m34s575)
- c. Ooda: daut doa jeforw e word.  
 or REL there painted got  
 ‘Or: “that got painted.”’  
 (F29, 2011-08-04 (01), 07m41s710–07m43s610)
- d. Dee Kaut saut besied de Däa, [fs]waut doa jrods,[/fs] [.] waut doa jeforw e woat.  
 the cat sat beside the door REL there just REL there painted got  
 ‘The cat sat beside the door that was just, that was getting painted.’  
 (M03, 2011-10-27 (01), 08m57s505–09m01s945)
- e. De Kaut sett bisied de Däa, // [fs]waut doa f-,[/fs] // waut jeforw e woat.  
 the cat sits beside the door REL there HES REL painted gets  
 ‘The cat is sitting beside the door that, that’s getting painted.’  
 (M17, 2011-10-29 (01), 10m17s965–10m22s905)

Besides the prompt sentence, what is common to all these examples is that the instances of variation in question all occur utterance-medially, in this case before some form of *woaren* ‘to become’. Reviewing other instances of this variation in the *Fibel* Corpus, it becomes apparent that this variation is rarely found in utterance-final position, although it appears in many other lexical and phonological environments, as the examples in (17) demonstrate:

- (17) a. Ons Bauss saigt, wi kjenne vondöag tiedig Meddag äten.  
 our boss says we can.PL today early lunch eat.INF  
 ‘Our boss says we can eat lunch early today.’ (S05)  
 (M02, 2011-08-03 (02), 01m44s835–01m48s395)
- b. Wan se de oole Koa vekofft ha'n, // dan woare se spazearen komen.  
 if they the old car sold have then will.PL they visit.INF come.INF  
 ‘If they’ve sold the old car, then they’ll come visit.’ (S27)  
 (M11, 2011-08-05 (01), 15m40s580–15m44s710)

Interestingly, although *-e* forms are overwhelmingly found in utterance-medial position, often before fricatives, this variant is not categorically present in this environment. Rather, it appears to be associated with connected, casual speech. In (18), contributor M16 offers two repetitions of his translation of S18 (“I’m always the one that has to weed the garden”), with the first

recorded as an informal ‘trial run’ before the main translation and the second representing one of the two (identical) primary translations. Only in the casual example do we find *-e* where the primary translations both have *-en*:

- (18) a. Ekj si emma dee jansja, waut däm Goad **e** weede saul.  
 I am always the one REL the.ACC garden weed.INF shall  
 ‘I’m always the one that’s supposed to weed the garden.’ (casual speech, *-e*)  
 (M16, 2011-07-23 (01), 19m53s400–19m56s290)
- b. Ekj si emma dee jansja, waut däm Goad **en** weeden saul.  
 I am always the one REL the.ACC garden weed.INF shall  
 ‘I’m always the one that’s supposed to weed the garden.’ (careful speech, *-en*)  
 (M16, 2011-07-23 (01), 20m00s120–20m03s690)

Utterance-medial variation in *-e(n)* among speakers in the central cluster thus differs markedly from the less predictable patterning of *-e(n)* among transitional speakers (where utterance-final occurrences are well attested) and appears to represent a form of reduction in casual, connected speech. Variation in *-e(n)* among these speakers is remarkable in several respects: not only is a uniform distribution almost always attributed to *-e(n)* forms across entire speech communities in linguistic studies of Mennonite Plautdietsch, but similar variation in *-e(n)* has never been reported in any related Mennonite groups. While this reduction-related variability may represent a later development specific to central Saskatchewan Valley Plautdietsch speakers, Mitzka (1968a [1922]: 203) notes frequent reduction of *-en* to *-e* before fricatives and /g/ in certain Polish varieties of Plautdietsch, suggesting that this phenomenon may have deeper historical roots. In light of this variability, the apparent social markedness of *-e(n)* variation in the Saskatchewan Valley is all the more interesting. Speakers from both the central and peripheral clusters commented on this feature as a significant dividing-line between groups within the region. Contributor M10, commenting on this distinction following his translation of S03 (“Children, can you pick her some strawberries and blueberries?”), approaches these differences with wry humour, separating the *-en* forms he associates with the ‘Old Colony’ speakers east of Highway 12 from his own *-e* forms among the ‘missionaries’ further west:

- (19) M10: Oba “plekjen” deed' wi nich.  
 CDC: “Plekjen” nich.  
 M10: Nä, [.] nä. Plekje. // Dee Ooltkol'nia von... // dee, woont aun... //  
 Ooste von 'em Numma 12,  
 CDC: *Uh-huh.*  
 M10: daut weare de Heide. Un aun de aundre Sied weare de Missionoare.  
 CDC: (laugh)  
 M10: Jo, un, un, [.] un wi wohnde met de [.] Missionoare.  
 CDC: (laugh)  
 M10: Doawääjen, uh, // ha' wi, uh, [.] woo nannt maun 'en *halo?* // (laugh) //  
 Jo, oba wi musste däm emma sea polische. // (laugh) //  
 [voice quality="laugh"]Dee word sea e'jejrint[/voice].
- M10: But we didn't (say) *plekjen* (to pick, -en).  
 CDC: Not *plekjen*.  
 M10: No, no. *Plekje* (-e). // The Old Colonists from... // the ones on... //  
 east of the Number 12 (highway),  
 CDC: *Uh-huh.*  
 M10: those were the heathens. And on the other side were the missionaries.  
 CDC: (laugh)  
 M10: Yes, and, and we lived with the missionaries.  
 CDC: (laugh)  
 M10: That's why, uh, // we had, uh, what do you call a *halo?* // (laugh) //  
 Yes, but we always had to really polish it. // (laugh) //  
 [voice quality="laugh"]It really got dirtied up.[/voice]  
 (2011-08-03 (01), 04m10s290–04m34s625)

Likewise, contributors M01 and F06 offer similar commentary on these divisions in the context of translating S13 (“Mom wanted to cook four big hams for supper today”). At the outset, M01, knowing that contributors to the *Fibel* represent communities throughout the Saskatchewan Valley, begins with a question about *-e(n)* variation, leading into a well-known rhyme that makes light of these differences. For his part, M01 is clear that this variation separates Old Colonists from non-Old Colonists, and that he and F06 are unambiguously part of the latter group—a point with which F06 agrees, but also elaborates on further below.<sup>55</sup>

---

55 Note that, unlike these differences in *-e(n)*, variation in pre-velar fronting (e.g., *koake(n)* vs. *köake(n)* ‘to cook’, etc.) encountered throughout (20) is not subject to any commentary. M01 attributes fronted pre-velar forms to both his and F06’s *Russländer* group and to the opposing ‘Old Colony’ speakers—despite the fact that he and F06 use entirely different variants of this feature.

- (20) a. M01: *How do you find your different people, // uh, with the German? // "Kö-, köaken, möaken, Schinkjenknöaken," // (laugh) ooda "köake, möake, Schinkjeknöake"?*  
 CDC: *(laugh) // Is that one of these differences? Like...*  
 M01: *That's Old Colony, and the... // never knew that?*  
 CDC: *Well, who says, y-, who says which one?*  
 M01: *We do the "koake, möake, Schinkjenknöaken."*  
 CDC: *Oh, okay.*  
 M01: *Old Colony do "köaken, möaken, Schinkjenknöaken."*  
 CDC: *Oh!*  
 M01: *Never knew that?*  
 F06: *Yeah, but you know, [.] th-, then there's all of the...*  
 CDC: *Na, wää saigt "köaken"?*  
 M01: *Wi doone "köake".*  
 CDC: *Oh, na, jo.*  
 M01: *See, that's the difference.*  
 F06: *"Foahre," un "foahren." (laugh)*  
 CDC: *Jo, yeah.*  
 F06: *"Foahren."*
- M01: *How do you find your different people, // uh, with the German? // "Kö-, köaken (to cook, -en), möaken (to make, -en), Schinkjenknöaken (ham bone, -en)," // (laugh) or "köake (to cook, -e), möake (to make, -e), Schinkjeknöake (ham bone, -e)"?*  
 CDC: *(laugh) // Is that one of these differences? Like...*  
 M01: *That's Old Colony, and the... // never knew that?*  
 CDC: *Well, who says, y-, who says which one?*  
 M01: *We do the "koake (to cook, -e), möake (to make, -e), Schinkjeknöaken (ham bone, -en)."*  
 CDC: *Oh, okay.*  
 M01: *Old Colony do "köaken (to cook, -en), möaken (to make, -en), Schinkjenknöaken (ham bone, -en)."*  
 CDC: *Oh!*  
 M01: *Never knew that?*  
 F06: *Yeah, but you know, th-, then there's all of the...*  
 CDC: *Well, who says "köaken" (to cook, -en)?*  
 M01: *We do "köake" (to cook, -e).*  
 CDC: *Oh, alright.*  
 M01: *See, that's the difference.*  
 F06: *"Foahre" (to drive, -e) and "foahren" (to drive, -en). (laugh)*  
 CDC: *Yes, yeah.*  
 F06: *"Foahren" (to drive, -en).*  
 (2012-10-18 (01), 13m38h555-14m14s025)

At this point, F06 turns the conversation back to an apparent exception to this pattern, noting variation in her extended family for this feature, despite their immediate kinship:

- (19) b. F06: *Like, my uncle (name), // uh, Uncle (name) on the farm,*  
 CDC: *Jo.*  
 F06: *dee fua-, uh, dee "foahren" uk emma.*  
 CDC: *Dee [voice quality="laugh"]foahren[/voice].*  
 F06: *Uh, "koaken." // Jo. // [fs]Un o-,[/fs] un, uh, // them, [.] her, uh, their mother and my mother were sisters, and we nev-, t-, never talked that way.*  
 CDC: *Well, yeah.*  
 F06: *Uh, it's strange.*  
 CDC: *Yeah.*  
 F06: *And Onkel (name) // was a brother, // (full name), // un dee deede uk nich soo räde. // They [.] talked like we did.*  
 CDC: *They did, jo?*  
 F06: *Dee [.] deede nich "foahren."*  
 M01: *They never...*  
 CDC: *Dee wearen Jeschwista, nich?*  
 F06: *Jo! Aula dree.*
- F06: *Like, my uncle (name), // uh, Uncle (name) on the farm,*  
 CDC: *Yeah.*  
 F06: *they, uh, they always drive (foahren, -en), too.*  
 CDC: *They [voice quality="laugh"]drive (foahren, -en)[/voice].*  
 F06: *Uh, cook (koaken, -en). // Yeah. // And, and, uh, // them, [.] her, uh, their mother and my mother were sisters, and we nev-, t-, never talked that way.*  
 CDC: *Well, yeah.*  
 F06: *Uh, it's strange.*  
 CDC: *Yeah.*  
 F06: *And Uncle (name) // was a brother // (full name), // and they also didn't talk that way. // They [.] talked like we did.*  
 CDC: *They did, yeah?*  
 F06: *They didn't drive (foahren, -en).*  
 M01: *They never...*  
 CDC: *They were all siblings, weren't they?*  
 F06: *Yes! All three of them.*  
 (14m14s355–14m41s585)

Comments such as these suggest a more complex situation with respect to *-e(n)* variation than is apparent from the distribution plots seen above. It is clear that variation in these features is not easily reduced to a single phonological or social explanation, or to uniform distributions that hold equally across all groups of speakers. While peripheral and Catholic speakers show essentially consistent use of *-e* in the present sample, the reflections of contributors such as F06 make clear that this may not always be the case, and that individual families may even demonstrate internal variation on this point. Despite the reported markedness of these variants among religious and geographical subgroups in the area, variation persists in previously



unreported ways, with members of the central cluster frequently showing reduction of *-en* to *-e* in utterance-medial contexts. Taken together, these observations suggest that existing generalizations concerning the distribution of *-e(n)* across Mennonite communities should perhaps be taken with a grain of salt. If exceptions to these patterns are well-attested in even the relatively limited sample of speech available in the *Fibel* Corpus, then it seems likely that similar exceptions might be anticipated in other, historically related speech communities, as well.

Although these morphological features are thus substantially more complex than the preceding phonemic-phonological and lexical ones considered earlier, it is worth noting that clustering on the basis of these variables produces similar groupings of speakers. With the exception of the merger of Catholic and peripheral Mennonite speakers into a single cluster (primarily due to their exclusive use of *-e* variants), the clusters that emerge under consideration of morphological differences broadly resemble those derived from lexical and phonological features, revealing larger groups of centrally located speakers and a smaller, geographically more disparate group of peripheral speakers. The recurrence of these clusters suggests that relatively coherent profiles of variation may exist for certain groups in the Saskatchewan Valley, although this remains to be confirmed in the syntactic and aggregate analyses below.

#### 5.2.4 *Syntactic clustering*

As Sections 1.2.2 and 4.2.1.5 noted, syntactic variation remains sparsely documented in previous research on Plautdietsch. Although some lexical-phrasal variables considered in Section 4.2.1.2 could be considered to have a syntactic dimension to them, even with these additional features included, only a relatively narrow range of syntactic variables in Plautdietsch speech communities have received analytical attention to date. Consequently, the *Fibel* Corpus focuses its primary attention on one area in which syntactic variation is known to exist in Mennonite Plautdietsch: in the linear order of elements in verbal complementation constructions (Kaufmann 2003b, 2005; Cox 2008, 2011a).

As more schematic entities than many of the constructions seen thus far, verbal complementation constructions require a different form of coding from other variables in the *Fibel* Corpus. As Section 4.2.1.5 details, coding for instances of these constructions made note of their observed complement orders (e.g., 1-2, where the matrix verb  $v_1$  precedes its complement

$v_2$ ; or 2-1, where the matrix verb follows its complement), transitivity (i.e., transitive, intransitive), the type of clause in which they appear (e.g., complement, conditional, etc.), and the actual complementation construction (e.g., perfectives, passives, modals, etc.). These features allow for investigation of several related phenomena at once. Not only do they permit consideration of complement ordering patterns that may be associated with particular constructional contexts or groups of speakers, but they also allow for investigation of more specific phenomena, such as so-called “verb-raising” and “verb projection raising” constructions (cf. Wurmbrand 2004). While these two labels make reference to specific, movement-based generative syntactic analyses that first drew sustained attention to these phenomena, both terms are now used to refer to these constructions without reference to any specific analytical framework. In verb-raising constructions, the matrix verb appears before its complement, but after any objects or modifiers of the complement verb, as in (21a); while in verb projection raising, objects and modifiers appear between an initial matrix verb and its complement, as in (21b):

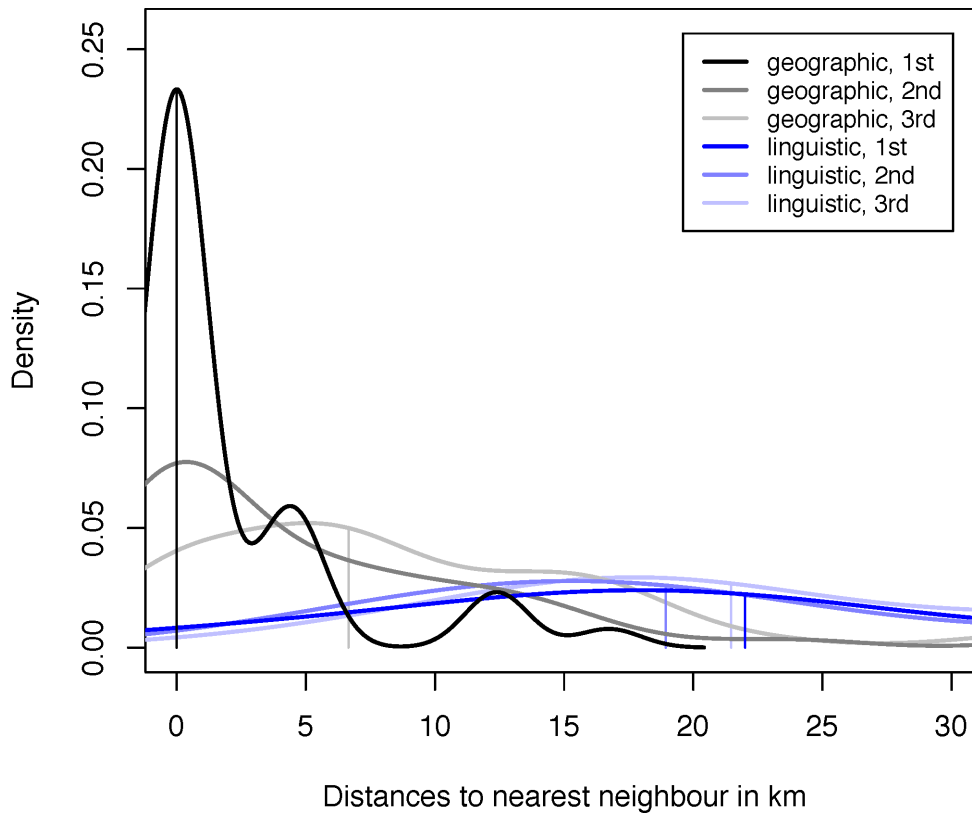
- (21) a. Wan wi Schnee **welle<sub>1</sub>** **sehne<sub>2</sub>**, // (...) mott' wi noh de Jebirje foahre.  
 if we snow want<sub>1</sub> see<sub>2</sub>:INF must we to the mountains drive: INF  
 ‘If we want to see snow, we have to drive to the mountains.’ (S38)  
 (M10, 2011-08-03 (01), 45m31s525–45m38s700)
- b. Ekj si frooh, daut wi // toop **kjännenn<sub>1</sub>** Beakja [.] **läsen<sub>2</sub>**.  
 I am happy that we together can<sub>1</sub> books read: INF<sub>2</sub>.  
 ‘I’m happy that we can read books together.’ (S12)  
 (F08, 2011-08-08 (01), 06m28s605–06m32s585)

Although these syntactic variables are narrower in scope than the other variables treated above, these few features are still sufficient for investigating several aspects of syntactic variation. Indeed, given the relative sparseness of many syntactic variables compared to morphological or phonological ones, investigating syntactic variation through carefully designed translation tasks such as this has not only been common practice in previous studies of Mennonite Plautdietsch, but has also been explicitly advocated as one means of achieving reliable coverage of relevant features (cf. Kaufmann 2005). The adoption of a similar approach here thus provides an opportunity to pursue two related goals at once, investigating both how syntactic variation patterns across the Saskatchewan Valley and how these methodological recommendations play

out in the case of the *Fibel* Corpus.

The syntactic phenomena under investigation here, limited as they are to a small number of complement orders, also differ from the remaining non-syntactic features in their treatment as continuous, rather than categorical, variables. Rather than treating complement orders as all-or-nothing, binary features, these syntactic variables count the number of occurrences of particular complement orders in the contexts of interest (e.g., the number of 1-2 orders attested overall, or in transitive vs. intransitive contexts, or in modal-infinitive constructions, etc.). Since different contributors may have provided different numbers of translations matching a particular syntactic frame—a sentence that was intended to elicit a modal construction may sometimes have been translated without the modal, for instance—counts of verbal constituent orders were normalized as percentages, allowing for more even comparison across speakers in each context of interest. The resulting normalized counts compare verbal complement orders overall (1-2 vs. 2-1, 1-2-3 vs. 1-3-2), as well as in particular constructional contexts (1-2 vs. 2-1 in MOD-INF, AUX<sub>PASS</sub>-PART, and AUX<sub>PERF</sub>-PART constructions), transitivity contexts (1-2 vs. 2-1 in intransitive and transitive contexts), and clause types (1-2 vs. 2-1 in complement, causal, conditional, and relative clauses), as well as the percentage of VR and VPR constructions noted in contexts where these could be observed.

Using these variables, it is possible to consider the overall relationship between physical geography and linguistic distance reflected in syntactic variation. In Figure 40, much the same flat density distribution appears with the syntactic variables as with the preceding phonemic-phonological, lexical, and morphological variables: contributors to the corpus are rarely most syntactically like their nearest geographical neighbours.



*Figure 40.* Density plot of distances to nearest geographical and linguistic neighbours, syntactic variables.

Even if geographical distance is not clearly related to syntactic differentiation, other trends may still be observable in these variables. Multidimensional scaling provides some evidence of structure in this syntactic variation, as seen in Figure 41. Here, contributors are arranged essentially along a single axis, with individuals such as M00 and M16 at one pole and others such as F01 at the other.

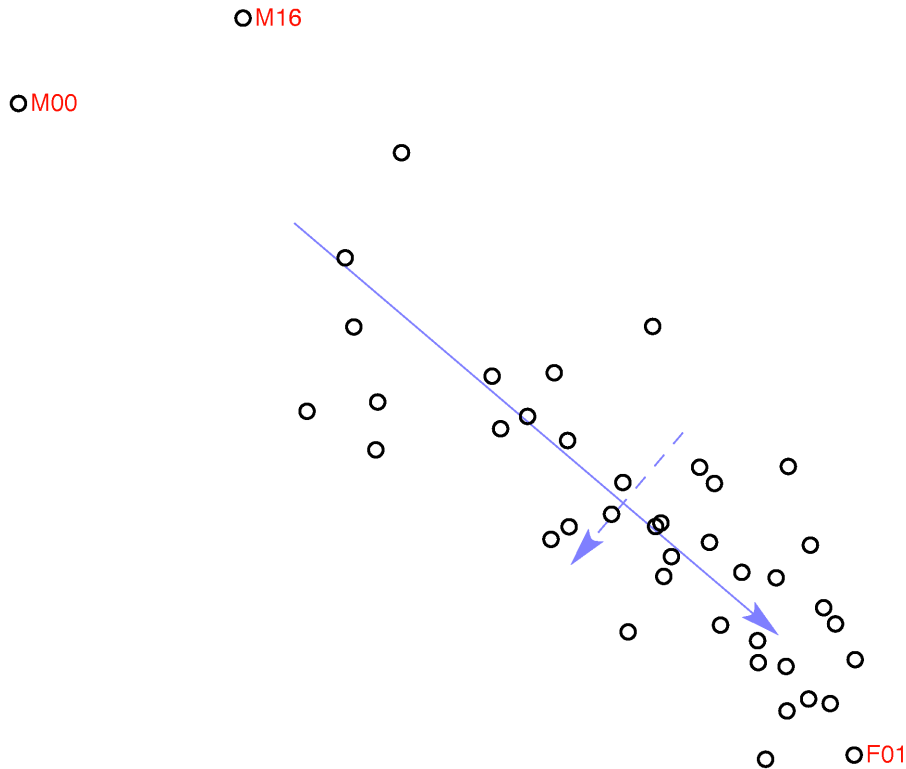
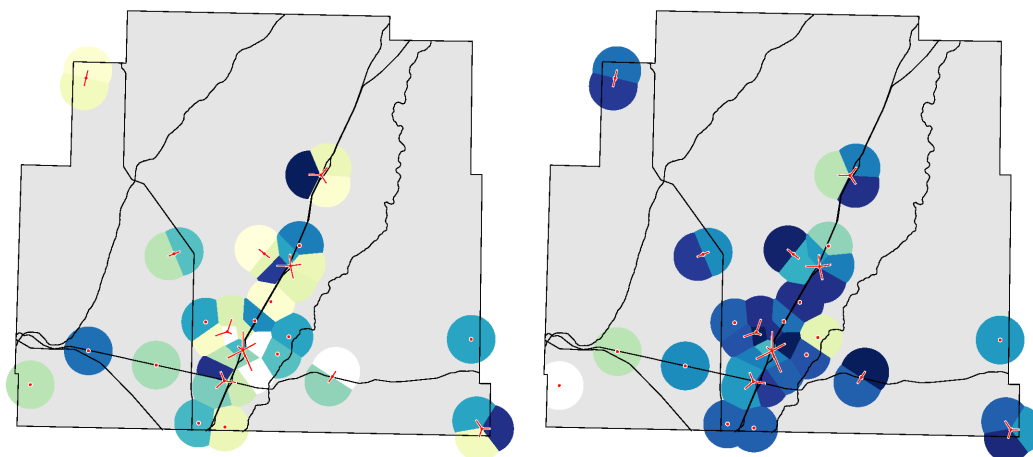


Figure 41. Two-dimensional MDS plot, syntactic variables ( $r = 0.98$ )

An inspection of the normalized responses for these speakers makes this visualization easier to interpret: contributors M00 and M16 both have quite low percentages of 1-2 complement orders in general (10% and 19%, respectively), while F01 has 1-2 complement order without exception in all responses. Not unexpectedly, then, given the pronounced weighting of these variables towards two-element constructions, the primary dimension of variation on this view can be related to 1-2 vs. 2-1 complement orders, with clear differences among speakers in their overall preference for one option over the other.

These responses also provide an opportunity to consider individual syntactic phenomena in greater detail. Figure 42 shows the distribution of the verb raising (VR) and verb projection raising (VPR) constructions introduced above, with occurrences of these constructions presented in blue. While VPR is shared by many speakers across the Saskatchewan Valley, VR is only attested with any real frequency among a handful of contributors, most prominently F23 (the uppermost, peripheral dark blue point in the plot), F26 (among the Humboldt Catholic contributors in the lower right-hand corner of the plot), and M15 and F28 (in the central cluster

of speakers). Notably, this distribution does not coincide with any of the groupings identified in the preceding sections. Speakers demonstrating higher-than-average proportions of VR constructions are members of all of the previously identified clusters, with no particularly strong association with one or another group.



*Figure 42.* Distribution plots of verb raising (left) and verb projection raising (right) constructions.

Having preliminary evidence of linguistic variation among speakers in syntactic constructions, it may be interesting to know more about how this variation is distributed among contributors and, in turn, how these groupings are arrayed across the Saskatchewan Valley. As before, clustering the contributors based on the available variables provides one way of approaching these questions. Applying agglomerative hierarchical clustering to these syntactic variables results in three groups in speakers, as seen in Figure 43 and Figure 44.

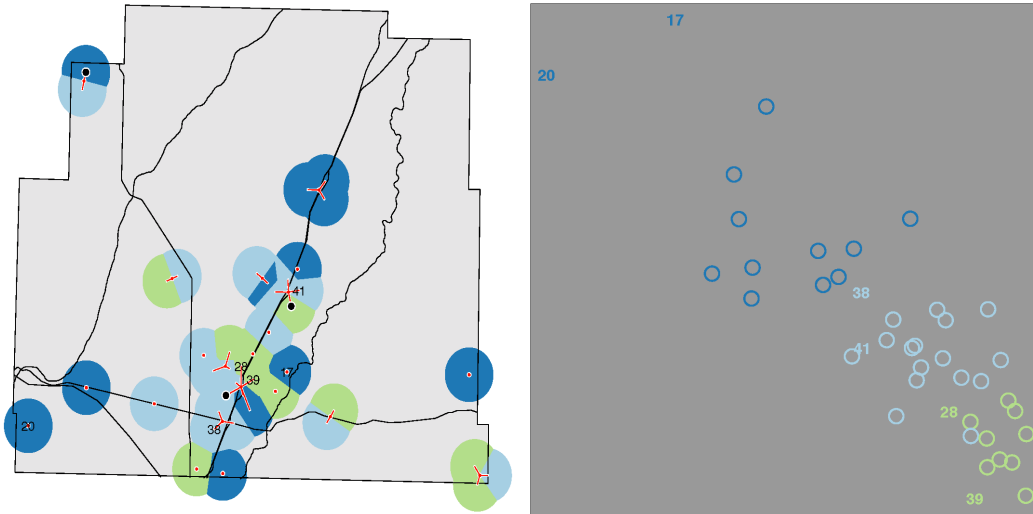


Figure 43. Hierarchical clustering by syntactic features (Ward's method, three clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.98$ ; right).

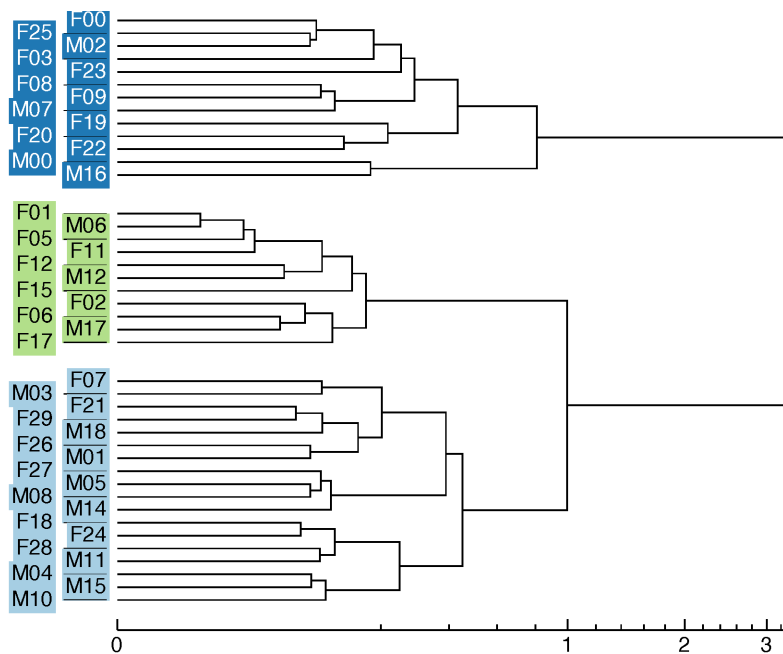
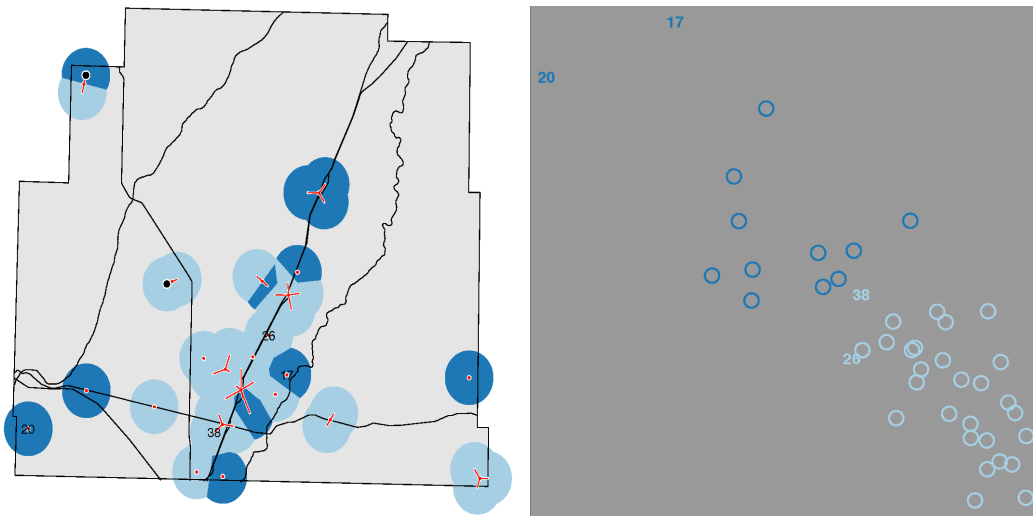


Figure 44. Dendrogram of hierarchical clustering by syntactic features (Ward's method, three clusters)

Although several features of these groupings deserve further attention, the corresponding two-dimensional MDS plot raises immediate questions as to the necessity of a three-way division among these speakers. In this plot, the clusters in light green and light blue share much the same space in the lower right-hand quadrant of the plot, with little definite separation between them.

By comparison, the same clustering procedures arrive at a much clearer break between dark and light blue clusters in Figure 45, where only two divisions are indicated. Here, the division between light blue and light green speakers collapses, leaving only a single, larger cluster situated in the lower right-hand quadrant of the MDS plot, a pattern that is reflected in the dendrogram in Figure 46, as well.



*Figure 45.* Hierarchical clustering by syntactic features (Ward's method, two clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.98$ ; right).



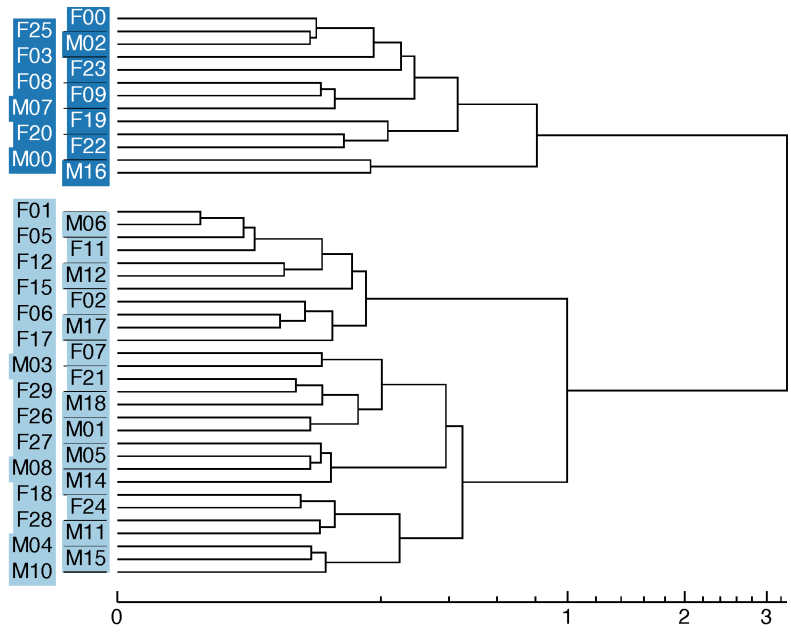


Figure 46. Dendrogram of hierarchical clustering by syntactic features (Ward's method, two clusters)

It thus appears justifiable to treat the light green and light blue clusters as one, given the low level of division evidenced between them. This conclusion is further supported by the distance measures associated with the highest-ranked cluster determinants for these groupings presented in Table 22 below. While both the light green and dark blue clusters show relatively high average distance measures, the intermediate light blue cluster is much less well defined, again suggesting that a merger between the light green and light blue clusters may be possible.

<i>Light green</i>	<i>Light blue</i>	<i>Dark blue</i>
AUXPERF-PART	INTR	COMP
1.64 (-1.24, 0.41)	0.85 (-0.79, 0.06)	1.54 (-0.68, 0.86)
COMP	<i>Overall</i>	INTR
1.36 (-1.13, 0.23)	0.84 (-0.80, 0.04)	(1.30, -0.50, 0.80)
REL	COMP	AUSPASS-PART
1.31 (-1.04, 0.27)	0.74 (-0.67, 0.07)	1.17 (-0.78, 0.39)
INTR	CAUS	<i>Overall</i>
1.28 (-0.99, 0.30)	0.64 (-0.61, 0.03)	0.96 (-0.21, 0.75)
<i>Overall</i>	MOD-INF	AUXPERF-PART
1.24 (-1.00, 0.25)	0.57 (-0.55, 0.02)	0.69 (-0.26, 0.42)

Table 22. Distance-based cluster determinants, syntactic variables.

Such observations about these clusters provides enough information to begin connecting these

more abstract results with the observed patterns of variation. From the dendrograms in Figure 44 and Figure 46, we note that the contributors who appeared at the extreme ends of the earlier MDS plot fall into different clusters here: M00 and M16 are both associated with the dark blue cluster, while F01 is treated as a member of the light green cluster. The preferences of these speakers for 2-1 and 1-2 complement orders, respectively, are also characteristic of the clusters to which they have been assigned: in general, contributors belonging to the dark blue cluster favour 2-1 orders, while speakers in the other clusters favour 1-2, with this being somewhat less pronounced among speakers in the light blue cluster.

This general association between clusters of speakers and particular constituent orders make the results of the clustering analysis more readily interpretable, and allows them to be related to what has been reported in previous research on these constructions. Among the above cluster determinants, the emergence of  $AUX_{PERF}-PART$  as the most significant determinant of the light green cluster (where 1-2 order is prevalent), is similar to a pattern reported by Cox (2008: 149), where 1-2 orders in these constructions were found to be particularly uncommon in some Saskatchewan Plautdietsch varieties. The consistency with which speakers in this cluster apply 1-2 constituent orders in this constructional context might be expected to emerge as a distinctive feature of this group, particularly if members of other clusters are less consistent in this respect. At the same time, however, clause type is not reported to be a significant predictor in the analysis in Cox (2011a), and is not addressed explicitly in Cox (2008), which leaves the trend towards 2-1 orders among speakers in the dark blue cluster in complement ( $COMP$ ) clauses less clearly explained.

Although some similarities thus exist between these clusters and what has been reported elsewhere on syntactic variation in Plautdietsch, in other respects, the preceding groupings diverge considerably from the results of clustering in the preceding sections. While Figure 43 and Figure 45 bear an intuitive resemblance to features of the preceding clusters—roughly, between the transitional-peripheral groups and the dark blue cluster, and the central group and the light blue and light green clusters—this similarity is likely not more than superficial. Only seven of the thirteen speakers in the dark blue cluster here are members of either the transitional or peripheral clusters in the preceding analyses, suggesting little reliable correspondence between these groups. Even if some association were to be demonstrated between the earlier peripheral-

transitional speakers and 2-1 orders, it would still leave unexplained why speakers such as M16, who pattern consistently with the central groups in all other respects, unexpectedly favour this particular complement order so markedly, or why the Humboldt Catholic contributors, who otherwise appeared to bear an overall closer similarity to the peripheral Mennonite speakers than any others, should now be divided between the light blue and light green clusters. In general, little of the geographical or sociodemographic coherence seen in preceding clustering analyses is evident here: rather, contributors who have, in all other respects, patterned together show considerable differentiation on these syntactic features.

With syntactically based clustering in this section differing markedly from other results, both in the groupings arrived at on the basis of other aspects of linguistic structure and in the results of previous investigations of syntactic variation in Saskatchewan Plautdietsch, it is possible that these differences can be attributed either to differences in population or in linguistic phenomenon. It may be that contributors to the present corpus are members of speech communities whose conventions of syntactic variation differ significantly from those of speakers in previous studies. Alternatively, it may be that variability in syntactic constructions such as these is substantially different in its own right from other forms of variation, and that the lack of correspondence between clustering in these constructions and those considered in the preceding sections is an accurate reflection of the overall character of syntactic variation in Saskatchewan Plautdietsch. Neither of these accounts necessarily precludes the other: speakers and communities at the centre of previous studies are not represented in the present corpus *and* that syntactic variation is inherently different from other forms of linguistic variation. Both offer intuitively plausible reasons for the observed differences and merit further consideration.

Several observations militate against these conclusions in the case of the *Fibel* Corpus. Considering possible differences between speaker populations first, it is unlikely that the Saskatchewan Mennonite Plautdietsch varieties considered in Cox (2008, 2011a) are exceptional among the forms of speech represented among Mennonites in the Saskatchewan Valley. Cox (2008) offers a defence of the typicality of these speakers, noting positive assessments of their language use on the part of other members of the local Mennonite community. Moreover, none of the speakers considered in these studies have sociodemographic profiles that diverge significantly from those of contributors to the *Fibel* Corpus, making it unlikely that their speech

communities are unrepresented in the present sample. On the whole, there is little evidence to suggest that the speakers in these studies are unrepresentative of their respective communities in their patterns of vernacular syntactic variation, or that they represent members of populations that are absent from the present corpus.

If these speakers are not atypical, then, and their respective speech communities have not been overlooked in the present study, the significant divergence between the patterns of constituent order variation observed in these speakers' non-elicited language use and those patterns observed among *Fibel* Corpus contributors calls for explanation. Comparing these previous studies with the syntactic responses offered by the *Fibel* Corpus, a pronounced skew becomes apparent in the present data that was not noted in earlier research. Even among members of the dark blue cluster, where 2-1 orders were found to be generally more prevalent, most speakers had 1-2 order in 40% or more of their responses, as is evident in the distribution plot in Figure 47 below. This is substantially higher than the average frequency of 1-2 complement orders reported by Cox (2008: 145) in non-elicited corpus data for a similar Saskatchewan Mennonite Plautdietsch speaker (14.9%).

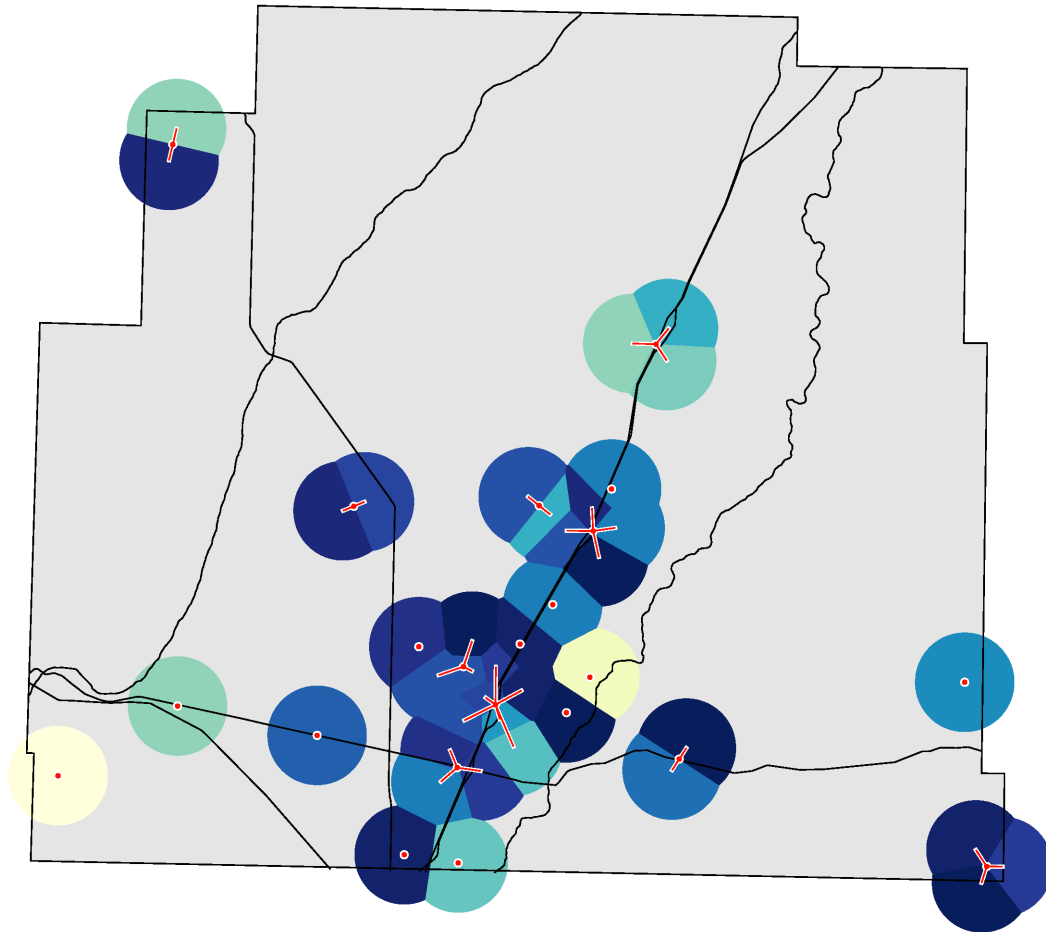


Figure 47. Distribution plot of verbal complement orders (1-2 in blue).

Such discrepancies between studies of the same speech communities based on elicited and non-elicited data raise suspicions that the syntactic responses in the *Fibel* Corpus data may not be entirely in keeping with vernacular conventions. Given the overall skew towards English-like 1-2 complement orders, task-related interference from the English sentences of the *Fibel* is likely to have exerted some influence on the corresponding translations. Indeed, this would not be unprecedented in studies of syntactic variation in other diasporic Mennonite Plautdietsch communities. Kaufmann (2005) reports comparable patterns of influence from prompt languages in a translation task-based study of variation in verbal complementation constructions in predominantly Latin American Mennonite Plautdietsch speech communities, lending further plausibility to the hypothesis of task-related interference. It is not unreasonable to suspect that bilingual speakers, when presented with a sentence in a closely related language for translation,

may favour the nearest structural equivalent of this prompt in the target language when this represents a licit option. This offers another possible explanation for the discrepancy between the results of syntactic and other forms of clustering on the basis of *Fibel* Corpus data: if task-related interference contributes to a general increase in the proportion of English-like complement orders in the corpus, then this, rather than any intrinsically different feature of syntactic responses, is likely to have contributed to the divergence between these clusters and those of other sections.

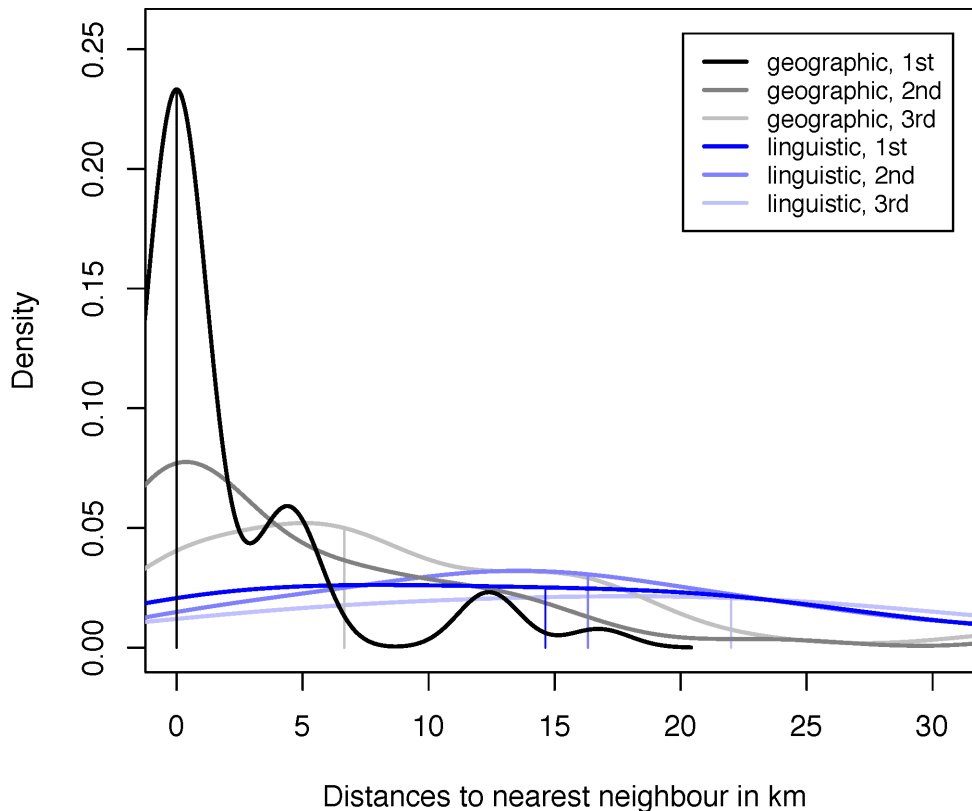
Taken together, these observations suggest significant differences between observations of syntactic variation in the *Fibel* Corpus and in samples of non-elicited discourse among members of Saskatchewan Plautdietsch-speaking communities. The observed skew towards English-like complement orders in the *Fibel* Corpus is consonant with the hypothesis that influence from the source language of the translation task underlying the corpus may be at play, generally favouring complement orders in the target language that more closely resemble their corresponding stimuli. This calls into question the adequacy of the present translation-task-based methodology in targeting vernacular syntactic variation, despite the arguments made by Kaufmann (2005) in favour of such methods in their ability to gather data on syntactic variation with sufficient consistency and in sufficient volume to allow for later quantitative analysis. While it is difficult to consider the present syntactic clustering wholly reliable, a valuable methodological lesson can nevertheless be drawn from this experience: translation tasks involving languages of similar typological profile may not always represent the ideal way of gathering information on vernacular syntactic variation, *pace* the recommendations of Kaufmann (2005), without some form of counterbalance being present to identify the degree and direction of skew. As a consequence, the results of syntactic clustering seen here will be treated as tenuous, requiring further attention in future research.

#### 5.2.5 *Aggregate clustering*

As has already been noted, one advantage of dialectometric methods lies in their ability to derive measures of linguistic distance between respondents based on large numbers of variables. In the preceding sections, this facility with multivariate data was applied specifically to variables associated with particular areas of linguistic organization, and proved useful in

learning more about the overall patterning of variation in these areas. The same methods can also be applied to the categorical variables seen thus far as a whole, bringing together all of these responses—phonemic-phonological, lexical, and morphological—to explore the larger picture of variation in the Saskatchewan Valley. In combining these distinct sets of variables into a single macro-analysis, the visualization and clustering methods seen earlier are able to assess the relative contribution of individual variables at each associated level of linguistic organization to the overall grouping of speakers. This aggregate approach affords a somewhat broader perspective on the attested variation, complementing the more focused analyses found in the preceding sections.

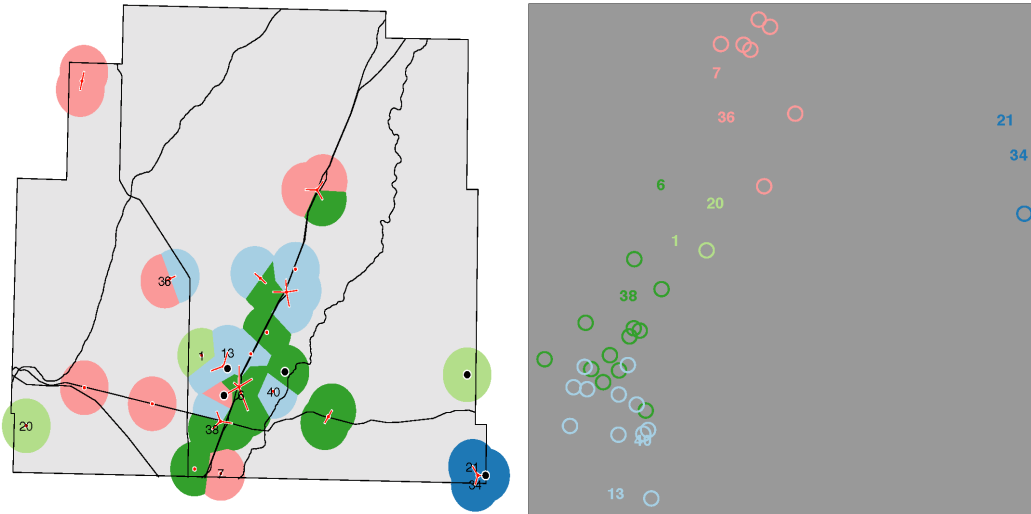
Combining the categorical variables from the preceding phonemic-phonological, lexical, and morphological analyses results in 7,194 instances of 167 variables being brought into this aggregate analysis. Regrettably, limitations in Gabmap prevent the inclusion of the syntactic variables in this analysis, whose coding produces continuous (‘count’) data, rather than the discrete (‘categorical’) forms associated with all other predictors. Even so, the remaining non-syntactic features are sufficient for arriving at a larger-scale view of variation in the Saskatchewan Valley (although they necessarily call for circumspection when interpreting the results of aggregate analysis, given the absence of variables representing other linguistic phenomena). To begin, it is worth noting that the lack of clear correlation between geographical distance and linguistic difference is just as apparent in Figure 48 under the aggregate analysis as it was in the preceding sections. On the whole, Plautdietsch speakers in the Saskatchewan Valley are rarely most linguistically similar to their nearest geographical neighbours, unlike in a traditional dialect continuum.



*Figure 48.* Density plot of distances to nearest geographical and linguistic neighbours, all categorical variables.

While geographical distance appears to be a poor predictor of linguistic similarity in general, this again does not imply that geographically more coherent patterns of variation cannot be identified for larger regions, or that speakers are entirely dissimilar to one another. This is evident in Figure 49, where hierarchical clustering on the aggregate set of features reproduces several of the linguistic divisions noted in preceding sections—most prominently, between peripheral and centrally located Mennonite speakers on the one hand (peripheral in light pink, central in light green and dark blue), and between Mennonite and Catholic speakers on the other (Mennonites on the left-hand side of the MDS plot below, Catholics in dark blue on the right).





*Figure 49.* Hierarchical clustering on all categorical features (Ward's method, five clusters; left) and corresponding two-dimensional MDS plot ( $r = 0.94$ ; right).

As the MDS plot suggests, some groups of speakers share a greater degree of linguistic similarity to one another than the geographically displayed clusters are able to indicate. The dendrogram in Figure 50 presents another view of these same clusters that makes their interrelation more apparent. In general, speakers in the centrally located clusters (in dark green and light blue) appear to be quite closely related, as are Mennonite speakers in the peripheral (light pink) and transitional (light green) clusters. Interestingly, the representatives of the Catholic Plautdietsch community do not form a branch separate from their Mennonite counterparts. On this analysis, it is clear that the Humboldt Catholic contributors have more of their linguistic characteristics in common with members of the peripheral and transitional Mennonite groups, and are not equally distinct from all Mennonite Plautdietsch speakers.

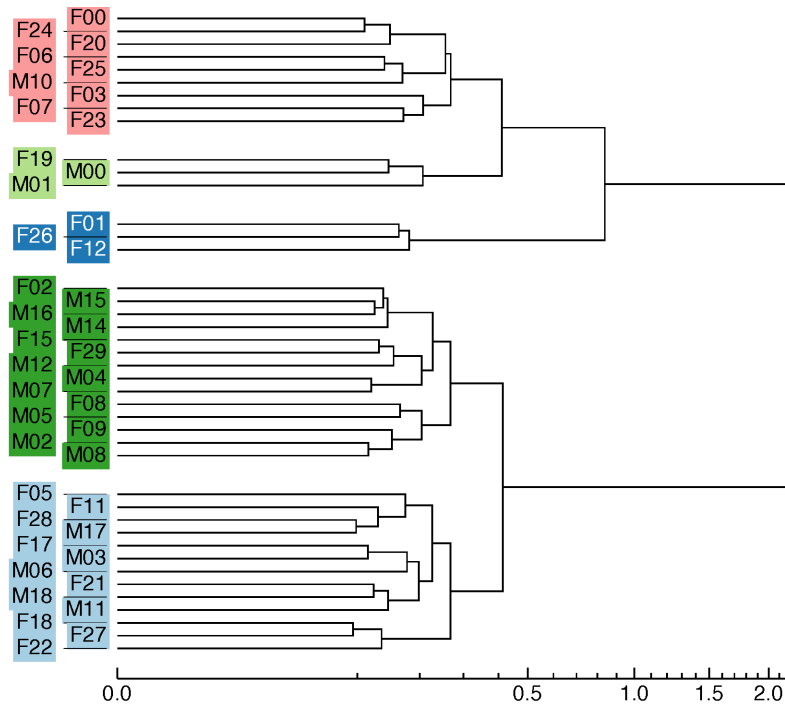


Figure 50. Dendrogram of hierarchical clustering on all categorical features (Ward's method, five clusters)

As before, more insight into the structure of these groupings can be gained from considering the linguistic features most closely associated with them. Table 23 and Table 24 present the highest-ranked cluster determinants for the two central clusters.

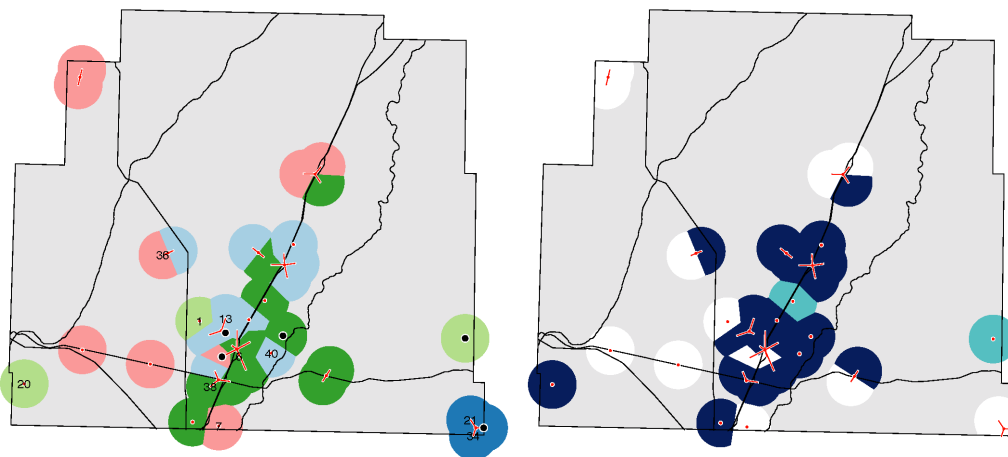
<i>Variable</i>	<i>Type</i>	<i>Distances</i>	<i>Associated Variants</i>
vUNTIL	LEX	1.45 (-1.03, 0.42)	<i>bott</i> (*bat)
vGIRLS	LEX	1.34 (-1.10, 0.24)	<i>Me(r)jalles</i> (*Me(r)jallen, Mäakjes)
vENGLISH	LEX	1.21 (-0.79, 0.42)	<i>Engelsch</i> (*Englisch)
vOEA	PHON	1.19 (-1.19, 0.00)	[eɔ] (*[øɔ], [oə], [oː])
vENNOUNSG	MORPH	1.14 (-1.03, 0.11)	<i>-en</i> (*-e)

Table 23. Aggregate cluster determinants, light blue (central) cluster.

<i>Variable</i>	<i>Type</i>	<i>Distances</i>	<i>Associated Variants</i>
vCOULD2SRREDUCTION	LEX-PHON	0.93 (-0.93, -0.01)	REDUCED (*FULL)
vTHETIMEXMSG	MORPH	0.81 (-0.79, 0.02)	<i>dän</i> (*däm, dee)
vTHATDAY	LEX-PHR	0.81 (-0.79, 0.02)	<i>dän Dag</i> (*däm/dee Dag)
vENNOUNSG	MORPH	0.80 (-0.74, 0.06)	<i>-en</i> (*-e)
vENNOMINAL	MORPH	0.80 (-0.74, 0.06)	<i>-en</i> (*-e)

Table 24. Aggregate cluster determinants, dark green (central) cluster.

Several observations can be made on the basis of this information. First, the overall distances associated with the determinants of the light blue cluster are significantly higher than those for the dark green cluster (Wilcoxon  $W = 25$ ,  $Z = 2.0226$ ,  $p = 0.0625$ ). This reflects the placement of the light blue cluster in the MDS plot above. On the whole, members of the light blue cluster show greater distinctiveness from other groups in their selection of variants than do members of the dark green cluster, hence their higher cluster determinant values here. Second, while the within-distance scores in each cluster are quite low, so too are their between-distance scores. That is, while speakers generally have conventions of variant selection that are similar to other members of their cluster for these top-ranked determinants, these variants are common to many members of other clusters, as well. Inspecting distribution plots for these cluster determinants suggests a possible reason for this trend: Figure 51 compares the aggregate hierarchical clustering seen earlier with the distribution of responses for vOEA. Considering the distribution of the variant [eq] across the central clusters, it becomes clear that both of the central clusters strongly favour use of this form. Similar sharing of characteristic variants is found across many of the highest-ranked cluster determinants for these two groups, suggesting that the closeness of these two clusters in particular might explain the low between-distance scores: as distinctive as these variants may be from those of other groups in the region, they are still most often common to both of the central clusters.



*Figure 51.* Hierarchical clustering on all categorical features (Ward's method, five clusters; left) and a distribution plot of vOEA (right, [eq] in blue).

Finally, the features summarized above give a clearer sense of the overall profile of linguistic variation in both of these two groups. In general, speakers in the central clusters appear to (a) favour *-en* endings over *-e*, a pattern most apparent in nominal constructions (vENNOUNSG, vENNOMINAL); (b) tend to use the *dän* variant of the masculine accusative singular definite article that derives from an earlier accusative form, rather than variants bearing a closer resemblance to historical dative (*däm*) or nominative (*dee*) forms (vTHETIMEXMSG, vTHATDAY); (c) have [eɔ] as a distinctive unrounded front realization of the phoneme <öa>, as seen in Figure 51 above; and (d) have certain lexical items that distinguish them from other speakers in the Saskatchewan Valley, such as *Mejalles* or *Mejallen* ‘girls’ (vGIRLS) and *Engelsch* ‘English’ (vENGLISH), where many other groups have *Mäakjes* and *Englisch*, respectively. Even under an aggregate analysis, characteristic linguistic features of each cluster such as these can still be determined, allowing for counterbalance between the abstraction that these methods favour and the descriptive detail needed to found such generalizations in specific patterns of observed variation.

A similar inspection of the determinants of the transitional and peripheral clusters in Table 25 and Table 26 brings to light several differences among members of these groups and the central clusters seen above, as well. While members of the transitional cluster share many of their variant selections with the central clusters, the transitional speakers show notable consistency in their use of *au* (vAUÉE, and more specifically vAUÉEATE), *some* (vSOME), and *daut* as the complementizer in neuter-headed relative clauses (vNEUTERRELCLAUSE) where the central clusters show greater internal variation. Unlike the variety of linguistic domains represented among the determinants of the central and transitional clusters, the highest-ranked cluster determinants of the peripheral group are predominantly phonological. With the exception of the *-e* ending in past participles (vENPASTPART), all other top determinants of the peripheral cluster are phonological or lexical-phonological in nature.

<i>Variable</i>	<i>Type</i>	<i>Distances</i>	<i>Associated Variants</i>
vSOME	LEX	1.72 (-1.61, 0.11)	<i>some</i> (*waut, walkje, ...)
vAuEe	LEX-PHON	1.69 (-1.17, 0.52)	<i>au</i> (*ee)
vAuEeATE	LEX-PHON	1.65 (-0.80, 0.85)	<i>au</i> (*ee)
vENDINGADJACCMSG	MORPH	1.60 (-1.03, 0.57)	<i>-a</i> (*-en)
vNEUTERRELCLAUSE	LEX-PHR	1.52 (-1.26, 0.26)	<i>daut</i> (*waut, woont)

Table 25. Aggregate cluster determinants, light green (transitional) cluster.

<i>Variable</i>	<i>Type</i>	<i>Distances</i>	<i>Associated Variants</i>
vOEA	PHON	2.02 (-1.19, 0.83)	[oə] (*[øɔ], [eɔ], [oː])
vREAL.OAPREVELAR	PHON	1.88 (-1.00, 0.88)	BACK (*FRONT)
vUE	PHON	1.80 (-0.51, 1.28)	[u] (*[y])
vENPASTPART	MORPH	1.76 (-1.19, 0.57)	<i>-e</i> (*-en)
vREAL.UAPREVELAR	PHON	1.72 (-0.99, 0.73)	BACK (*FRONT)

Table 26. Aggregate cluster determinants, light pink (peripheral) cluster.

These cluster determinants present distinctive phonological characteristics of the peripheral group quite clearly: members of this cluster tend to have back realizations of the /oa/ and /ua/ diphthongs in pre-velar environments, as well as for the phoneme /u/. That these phonological features might present a dividing-line between Mennonite groups in the Saskatchewan Valley is also suggested by M08, who attributes back-vowel realizations of pre-velar /oa/ to *Russländer* Mennonites in his comments on translating S36 ('The farmer's wife wanted to be sure that the water got boiled first'):

- (22) M08: Russlândia wudden saijen "koakt."  
 Russians would say boiled  
 'Russländer Mennonites would say *koakt* ('boiled', [oə]).'  
 CDC: Koa-...  
 HES  
 'Coo-...'  
 M08: Wo-, g-, de't Wota [emph]jekoakt[/emph] haud.  
 HES HES the water boiled:PTCP had  
 'The water had *boiled* ([oə])'  
 (2011-09-13 (02), 18m27s460–18m32s360)

As we have already seen, these are clearly not the only features that set this peripheral group apart from other Plautdietsch speakers in the Saskatchewan Valley. The cluster determinants of the central and peripheral groups also suggest ways in which Plautdietsch speakers in the region

vary from one another, and thus implicitly highlight other differences between those groups and peripherally located speakers.

Lastly, Table 27 presents the highest-ranked cluster determinants and their associated variants for the dark blue cluster of Catholic Plautdietsch speakers from the Humboldt area. As in the preceding phenomenon-specific analyses, marked differences are apparent between these speakers and other contributors to the *Fibel* Corpus. The distance measures presented here underscore both the internal coherence of the Catholic group in the use of these features (as reflected in the low within-distance scores), as well as the significant differences between Catholic and Mennonite speakers on all of these measures (as reflected in the high between-distance scores), particularly in the realization of the diphthongs /ea/ (vWASVOWEL, vHORSESVOWEL) and /oo/ (vOo) and several lexical items (vHOUSE, vUNCLES). While useful, these top-ranked determinants fail to draw attention to features common to this cluster and others in the present sample. Inspection of cluster determinants with lower between-distance scores, such as vENINF (within distance -1.24, between distance 0.30, total score 1.54) in Figure 52, provides evidence for the similarities suggested by the dendrogram in Figure 50 between the Humboldt Catholic cluster and peripheral Mennonite speakers.

<i>Variable</i>	<i>Type</i>	<i>Distances</i>	<i>Associated Variants</i>
vHOUSE	LEX	3.41 (-0.32, 3.08)	<i>Kot</i> (* <i>Huus</i> )
vWASVOWEL	LEX-PHON	2.92 (-0.40, 2.52)	<i>äa</i> (* <i>ia</i> )
vUNCLES	LEX	2.92 (-0.40, 2.52)	<i>Oohms</i> (* <i>Onkels</i> )
vOo	PHON	2.92 (-0.40, 2.52)	[o·] (*[əo])
vHORSESVOWEL	LEX-PHON	2.92 (-0.40, 2.52)	<i>äa</i> (* <i>ia</i> )

Table 27. Aggregate cluster determinants, dark blue (Humboldt Catholic) cluster.

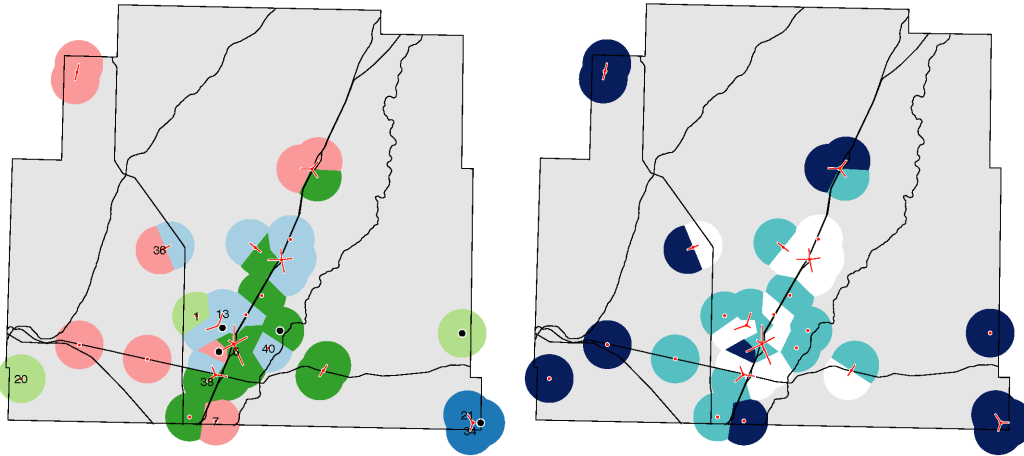


Figure 52. Hierarchical clustering on all categorical features (Ward's method, five clusters; left) and a distribution plot of  $vEN_{INF}$  (right,  $-e$  in blue).

This aggregational approach to exploring linguistic variation across all of the available categorical variables is thus useful for assessing the relative importance of different aspects of variation to the overall divisions that have been identified thus far. While most clusters of speakers appear to be distinguished most prominently by features from a range of areas of linguistic organization, others, such as the peripheral speakers in the corpus, are separated most immediately by marked phonological differences. Aggregate analysis also lends further support to the conclusion that linguistic variation in the Saskatchewan Valley is not distributed geographically in such a way that speakers are most similar to their nearest neighbours. While much the same broad, regional trends are suggested in the aggregate analysis as in the preceding, phenomenon-specific analyses, these groupings should be treated with some caution, as it has not been possible with these dialectometric methods to assess the possible contributions of sociohistorical or demographic factors to these distributions at anything more than a cursory level. Even so, the emergence of relatively coherent clusters of speakers that recur in both phenomenon-specific and aggregate analyses is encouraging, and provides a useful basis for further investigation.

### 5.3 Conclusions

The dialectometric methods seen in the preceding sections are able to offer considerable insight into the geographical dimension of linguistic variation in the Saskatchewan Valley. Not

only are they compatible with the kinds of documentation-based corpus data available to this study, but the breadth of exploratory methods that dialectometric tools such as Gabmap provide are particularly useful in arriving at an initial view of variation in the Saskatchewan Valley, where relatively little was known about the linguistic features of these communities in advance. These methods draw attention to six notable features of linguistic variation in the region in particular:

1. *Catholic-Mennonite differences.* A deep linguistic division appears to run between the Plautdietsch varieties of Mennonites in the Saskatchewan Valley and those of Catholics in the former St. Peter's Colony. While this might have been anticipated from the distinct histories of both groups, it is notable that this differentiation emerges without any sociodemographic having been provided to the inductive methods applied here. This division is not only apparent in the visualizations provided by dialectometric analysis, but its specific linguistic correlates are open to further study through attendant methods provided by tools such as Gabmap (e.g., through distribution plots of individual predictors, whether selected on the basis of theoretical interest or through the quantitatively derived cluster determinant measures seen in the preceding sections), thus refining our understanding of what specific linguistic features distinguish these two groups and which areas of difference may call for additional documentation.

2. *Regional differences.* Broad, regional differentiation is apparent in the *Fibel* Corpus data. Under both the phenomenon-specific and aggregate analyses, clusters of speakers with similar patterns of variant usage emerge in certain areas of the Saskatchewan Valley. These clusters divide the Mennonite contributors into two main areal groups: one in the central valley region, located near the geographical mid-point of the Saskatchewan Valley area and extending along the South Saskatchewan River; and another in the periphery of the valley, towards the northern and western edges of the Saskatchewan Valley closer to the North Saskatchewan River. (The transitional group identified under these analyses could also be considered to have some regional affiliation, being distributed from east to west across the central valley. With many members of this cluster having been raised outside of the Saskatchewan Valley proper, however, the grounding of this cluster in the geography of the region remains open to question; see below).

These observations of apparent regional differences immediately raise the following question: why would the Saskatchewan Valley, an area with a relatively recent history of



Mennonite settlement, be divided into geographical-linguistic regions at all? While the preceding dialectometric methods provide little assistance on their own in determining the relevance of non-geographical factors to these divisions, considering the historical demographics of the region reviewed in Chapter 2 proves instructive. Placing these linguistic divisions in the larger context of the history of Mennonite settlement in the Saskatchewan Valley, it becomes apparent that the central cluster situated around the South Saskatchewan River falls largely within the core Old Colony and Bergthaler settlement region, demarcated by the historical boundaries of the Hague-Osler Mennonite Reserve. Similarly, the peripheral cluster in the northern and western regions of the valley encompasses areas more densely settled by *Russländer* immigrants after the mass emigration of Old Colony and Bergthaler Mennonites to Latin America in the 1920s. These apparent geographical clusters are thus likely not entirely divorced from the historical events that shaped the Mennonite population of the region. As contributor M10 commented earlier, the peripheral Mennonite communities west of Highway 12 were not only linguistically distinct from those in the central Saskatchewan Valley, but also often differed in their overall religious orientation. Although the dialectometric methods used in this section draw attention primarily to the geographical face of this variation, additional sociodemographic correlates seem altogether plausible and will be considered more closely below.

3. *Geographical and linguistic distance.* Despite these pronounced regional trends in linguistic variation, there is little discernible relationship between immediate geographical proximity and linguistic similarity in any aspect of linguistic organization. While speakers may share a broadly similar set of linguistic variants with other speakers in their surrounding area, they are only rarely most like their nearest geographical neighbour in their linguistic conventions.

At first blush, this lack of a dialect-continuum-like, positive correlation between linguistic and geographical distance appears easily attributed to obvious demographic and historical factors, and thus less deserving of note. The relatively recent history of settlement in the region, coupled with the repeated dispersal of individuals of different linguistic backgrounds throughout the region (as a consequence both of migration out of the Saskatchewan Valley and of general agricultural expansion) and the weakening of once-prominent societal divisions between

the longer-resident *Kanadier* and the more recently arrived *Russländer* groups, makes it less than surprising that geographical neighbours may differ substantially in their linguistic preferences, despite their present-day physical proximity.

While these features may contribute to the overall heterogeneity of the local Mennonite community at a societal level, variation as reflected at the level of the individual also deserves attention. It is noteworthy that the several sets of siblings represented among the contributors to the *Fibel* Corpus are rarely linguistically most similar to one another. Despite their shared place of birth, upbringing, denominational background, and immediate sphere of close kinship relations, as well as the relatively minor differences in age between most such pairs, these siblings rarely show the same profiles of variant selection, other than maintaining membership in the same macro-groupings (e.g., central-*Kanadier* vs. peripheral-*Russländer*). Figure 53 reproduces the aggregate hierarchical cluster analysis from Section 5.2.5, with the clustering of pairs of siblings under this analysis given in Table 28.

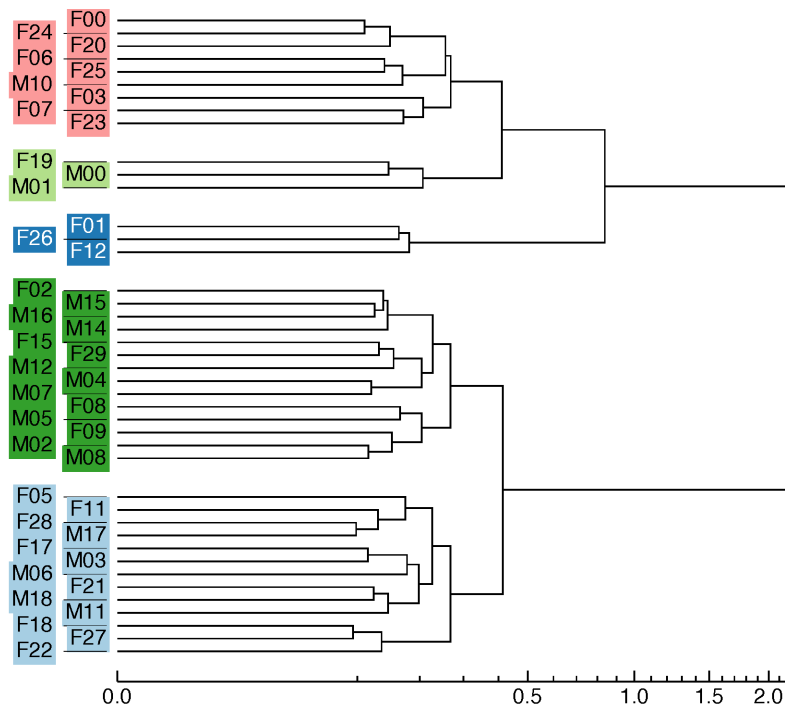


Figure 53. Dendrogram of hierarchical clustering by all categorical features (Ward's method, five clusters).

<i>Sibling A</i>	<i>Sibling B</i>	<i>Distance, A→B</i>	<i>Distance, B→A</i>
F09 (green)	F15 (green)	2 divisions (4 speakers)	4 divisions (8 speakers)
F18 (blue)	F27 (blue)	1 division (0 speakers)	1 division (0 speakers)
F21 (blue)	M07 (green)	6 divisions (12 speakers)	5 divisions (13 speakers)
M04 (green)	M05 (green)	4 divisions (8 speakers)	3 divisions (4 speakers)
M06 (blue)	M18 (blue)	2 divisions (2 speakers)	3 divisions (2 speakers)

*Table 28.* Sibling contributors and their clusters under aggregate hierarchical cluster analysis. Distances between sibling pairs are given in the number of intervening binary divisions and speakers.

The apparent linguistic distance between all but one set of siblings in their selection of variants is intriguing and not easily attributed to the differences in physical geography on which dialectometric methods typically concentrate. These differences do not appear to be coincidental: while a Wilcoxon signed rank test shows that the mean linguistic distance between pairs of siblings is significantly lower than between pairs of non-siblings overall (based on the aggregate distances computed by Gabmap over all categorical variables;  $W = 0$ ,  $Z = -2.8031$ ,  $p = 0.0001$ ), the same statistical test shows that there is no significant difference in mean linguistic distance between pairs of siblings and pairs of non-siblings who belong to the same cluster ( $W = 42$ ,  $Z = -0.8664$ ,  $p = 0.5787$ ). That is, on average, siblings are no more linguistically similar to one another than they are to non-siblings in the same cluster. This less obvious feature of variation in Saskatchewan Valley communities might be more readily understood through analytical methods with a greater capacity for incorporating non-geographical factors and when considered in the light of recent research on language variation in smaller, socioeconomically uniform communities characterized by dense, multiplex networks of interaction. These points are pursued further in the next chapter.

*4. Transitional speakers.* Dialectometric methods also identify the presence of ‘transitional’ speakers whose linguistic practices are not easily attributed to one or another group in the Saskatchewan Valley. Speakers in this transitional cluster tend to be geographically disparate, unlike all other major clusters, which have more obvious central geographical tendencies (see above). These speakers are also linguistically distinct, albeit in an interesting way: while sharing their repertoire of linguistic variants with other groups in the Saskatchewan Valley (and thus not separated by hard-and-fast isoglosses or exclusive, shibboleth-like use of particular variants), members of the transitional cluster show patterns of variant use that are not

attested in other Saskatchewan Valley groups. Where members of other clusters often show all-or-nothing distributions of certain variants that set them apart from other speakers (e.g., *-e* in one group, *-en* in another), members of the transitional group commonly alternate between many of the generally available variants.

The attested distributions of variants among these speakers is remarkable not only in the context of the Saskatchewan Valley, but also in reports on linguistic variation in the Russian Mennonite diaspora more generally. As summarized in Chapter 3, intra-speaker variation in many of the variables thought to characterize differences between varieties of Plautdietsch is not robustly attested: features such as *-e* vs. *-e(n)* in their many constructional permutations are virtually always treated as dividing-lines between whole speech communities, rather than as loci of linguistic variation at the level of individual speakers. In the Saskatchewan Valley, ‘transitional’ speakers present a clear exception to this generalization, presenting difficult evidence for descriptions of such features that assume an absolute, all-or-nothing distribution of such variants across speech communities, as is typical in much linguistic literature on Mennonite Plautdietsch.

5. *Stable divisions.* Separate analyses of variation in different aspects of linguistic organization support largely the same set of divisions throughout. Several separate clustering analyses, whether conducted on specific areas of linguistic variation or on all categorically treated variation as a whole, arrive at broadly similar classes of speakers (e.g., central, peripheral, transitional, Catholic), with more or less differentiation between these groups appearing in some areas of linguistic organization. These results, while clearly provisional, suggest some success on the part of these methods to identify coherent groups of speakers, as was identified as a goal of analysis at the outset of this chapter.

6. *Stable membership.* Relatedly, for the majority of speakers represented in the corpus, membership in the aforementioned clusters does not differ significantly from one linguistic phenomenon to another. Although several notable exceptions to this generalization exist that will be discussed below, most contributors to the *Fibel* Corpus demonstrate consistent affiliation with the same profile of variant selection across all linguistic domains.

This statement holds for all areas of linguistic organization except for variation in word order, where the *Fibel* Corpus does not produce reliable classifications. Here, the

methodological recommendations found in the literature on Mennonite Plautdietsch (e.g., Kaufmann 2005) for the balanced observation of syntactic variation through careful, translation-task-based elicitation have not produced the desired results. Although it would be entirely reasonable to believe *a priori* that speakers may not vary significantly in their word order preferences and that these results simply reflect that lack of differentiation in syntactic phenomena, preceding studies of Mennonite Plautdietsch speech communities in Saskatchewan (e.g., Cox 2008, 2011a) and elsewhere in the Russian Mennonite diaspora (e.g., Kaufmann 2003b, 2005, 2007, 2008) find significant differences between groups of speakers, calling into question the adequacy of the syntactic results in the corpus. The overall tendency towards English-like word orders—markedly more pronounced here than in previous studies of Saskatchewan Plautdietsch based on unelicited samples of language use—suggest possible task-related interference from the English-language prompts, although further research would be required to determine the degree to which these results diverge from more typical patterns of use.

While the remaining, non-syntactic classifications appear relatively stable, and while most speakers are consistently affiliated with the same classification for all non-syntactic aspects of variation considered here, there are still a number of speakers who demonstrate notable variability in their membership in these classes. This merits further consideration: why would some contributors to the corpus be associated with different profiles of variation for different areas of linguistic organization when most others are not? While other forms of analysis might be used to bring attention to sociodemographic factors that have not been included thus far, even a brief summary of the dialectometric classifications seen above, supplemented with several basic pieces of demographic information about the speakers themselves, sheds considerable light on this problem. Table 29 presents the results of the preceding clustering analyses for each contributor to the *Fibel* Corpus, adding to these results information about each speaker's place of birth and family background that was provided earlier in Table 17. (All colour coding in this table is facultative, and is intended only to make consistencies and discrepancies between clustering analyses more readily apparent)

<i>ID</i>	<i>Group</i>	<i>POB</i>	<i>Phonological</i>	<i>Lexical</i>	<i>Morphological</i>	<i>Aggregate</i>
F00	U	O	Peripheral	Peripheral	Peripheral	Peripheral
F01	H	H	Catholic	Catholic	Catholic-Periph.	Catholic
F02	C	C	Central	Central	Central	Central
F03	U	P	Peripheral	Peripheral	Peripheral	Peripheral
F05	C	C	Central	Central	Transitional <sup>†</sup>	Central
F06	U	P	Peripheral	Peripheral	Peripheral	Peripheral
F07	U	C	Peripheral	Peripheral	Transitional	Peripheral
F08	U	C	Peripheral	Peripheral	Central	Central
F09	C-U	C	Transitional	Central	Central	Central
F11	C	C	Central	Central	Central	Central
F12	H	H	Catholic	Catholic	Catholic-Periph.	Catholic
F15	C-U	C	Central	Central	Central	Central
F17	C	C	Central	Central	Central	Central
F18	C	C	Central	Central	Central	Central
F19	C	O*	Transitional	Central	Transitional	Transitional
F20	U	C/P	Peripheral	Peripheral	Peripheral	Peripheral
F21	C	C	Central	Central	Central	Central
F22	C	C	Central	Central	Central	Central
F23	U	C/P	Peripheral	Peripheral	Transitional <sup>†</sup>	Peripheral
F24	U	O	Peripheral	Peripheral	Peripheral	Peripheral
F25	U	O	Peripheral	Peripheral	Peripheral	Peripheral
F26	H	H	Catholic	Catholic	Catholic-Periph.	Catholic
F27	C	C	Central	Central	Central	Central
F28	C	C	Central	Central	Central	Central
F29	C	C	Central	Central	Central	Central
M00	C	O*	Transitional	Central	Transitional	Transitional
M01	U	C	Transitional	Peripheral	Central	Transitional
M02	C-U	C/P	Transitional	Central	Central	Central
M03	C	C	Central	Central	Central	Central
M04	U	C	Central	Central	Central	Central
M05	U	C	Central	Peripheral	Transitional	Central
M06	C	C	Central	Central	Central	Central
M07	C	C	Central	Central	Central	Central
M08	C	C	Transitional <sup>†</sup>	Central	Central	Central
M10	U	P	Peripheral	Peripheral	Peripheral	Peripheral
M11	C	(C)	Central	Central	Central	Central
M12	C	(C)	Central	Central	Central	Central
M14	C	C	Central	Central	Central	Central
M15	C	C	Central	Central	Central	Central
M16	C	C	Central	Central	Central	Central
M17	C	C	Central	Central	Central	Central
M18	C	C	Central	Central	Central	Central

Table 29. Summary of clustering analyses by speaker. (Group labels: C = *Kanadier*, U = *Russländer*/USA Molochnaya, C-U = mixed *Kanadier-Russländer*, H = Humboldt Catholic. POB labels: C = central Sask. Valley, P = peripheral Sask. Valley, O = outside of Sask. Valley, H = Humboldt area)

This summary makes apparent several features of the preceding analyses and their relation to other demographic categories. Perhaps most immediately noticeable here is the degree of alignment between demographic features related to speaker group (e.g., *Russländer* or *Kanadier* Mennonites, Humboldt-area Catholics, etc.) and place of birth (e.g., peripheral, central, transitional, etc.) and the results of clustering, resulting in rows of all the same colour. At the same time, other speakers show much less consistent patterning along these lines, with their respective rows showing a mix of colours. A closer inspection of both the more consistent and more varied classifications suggests several possible reasons for these differences. Those speakers for whom the results of clustering are more consistent fall into three classes:

1. Speakers from Humboldt Catholic families (F01, F12, F26);
2. Speakers from *Russländer* families who were raised either in a peripheral Saskatchewan Valley community or elsewhere in Saskatchewan (F00, F03, F06, F20, F23, F24, F25, M10);
3. Speakers from *Kanadier* families who were raised in a central Saskatchewan Valley community (F02, F05, F11, F17, F18, F21, F22, F27, F28, F29, M03, M06, M07, M08, M11, M12, M14, M15, M16, M17, M18).<sup>56</sup>

Likewise, speakers who show mixed results under these clustering analyses can also be assigned to three classes:

1. Speakers raised outside of Saskatchewan (F19, M00, both classified as transitional in all but their lexical features);
2. Speakers from *Russländer* families who were raised in the central Saskatchewan Valley (F07, F08, M01, M05, but not M04);
3. Speakers from ‘mixed’ families where one parent was of *Kanadier* background and the other of *Russländer* background (F09, M02, but not F15).

Differences between these two sets of speakers are thus not haphazard: the speakers with consistent linguistic classifications are, by and large, also consistent with respect to their family background and places of birth, while those speakers with less consistent linguistic

---

<sup>56</sup> Both F05 and M08 appear to have been misclassified as transitional in their morphological and phonological features, respectively. Closer inspection of the relevant MDS plots for both speakers suggest closer affiliation with the central cluster, with whom they are classified under all other analyses.

classifications are less homogenous in either or both of these respects.<sup>57</sup> The latter, more linguistically and demographically varied groups are of immediate relevance to the picture of linguistic variation in the Saskatchewan Valley being developed here. The observation of distinct patterns of variation among Mennonite Plautdietsch speakers from outside of Saskatchewan may suggest that these other Plautdietsch speech communities have differing conventions in their selection of variants, though this would clearly require further investigation to confirm. More immediately, however, the observation of linguistic variation among speakers of mixed *Kanadier-Russländer* background is interesting, especially given reports in other communities of ‘mixed’ varieties of Mennonite Plautdietsch emerging through the intermarriage of parents of different dialect backgrounds (cf. Mitzka 1930). Similarly, the variation encountered among contributors of *Russländer* background who grew up in predominantly non-*Russländer* communities in the central Saskatchewan Valley suggests that these speakers have (at least partially) adopted the norms of other, local Mennonite Plautdietsch speakers, although this too would require additional investigation to determine the extent and nature of this apparent linguistic convergence. For now, it is enough to note that socially intense contact between speakers of different varieties of Plautdietsch in the Saskatchewan Valley, while not observed among the majority of contributors, is well attested, and appears to have left its mark on the linguistic practices of at least some speakers in the area.

Several features of dialectometric analysis can be drawn on to inform the direction of future documentation efforts, pointing to areas in which further information may be particularly beneficial. At an intuitive level, the results of the clustering and multi-dimensional scaling analyses in the preceding sections immediately suggest areas where more documentary attention may be needed. In Section 5.2.1, for instance, it was noted that contributor F23, a Mennonite Plautdietsch speaker, most closely resembled the Catholic Plautdietsch contributors from the Humboldt area in her phonological features. While such phonological differences between F23 and other Mennonite speakers may simply have been coincidental, when one considers the personal information given in Table 17, it becomes apparent that F23 is exceptional in another

---

<sup>57</sup> That there should remain so much variation among the former groups who are characterized by such considerable demographic coherence is noteworthy in its own right, and will be discussed further in the following chapter.



way: F23 is the sole contributor to the *Fibel* Corpus whose parents are from the small group of Russian Mennonites who emigrated to the Saskatchewan Valley from the northern United States around the turn of the twentieth century. Although far from conclusive, this observation raises the possibility that speakers of this background (i.e., Molochnaya-descended Mennonites, primarily from Mennonite Brethren and Rosenorter / General Conference denominations, who emigrated to North America in the 1870s and 1880s; see §2.4) may differ linguistically from other Mennonites in the Saskatchewan Valley, and potentially bear a closer resemblance in at least some of their linguistic features to Catholic Plautdietsch speakers. The ability of dialectometric methods to bring attention to situations such as this is valuable, as it allows for further, focused attention to be given to smaller groups within the larger Mennonite community who might otherwise be overlooked.

The results of these clustering analyses can also be drawn on in another way to provide direction for ongoing documentation. In the preceding sections, several distance metrics were applied to clusters of speakers as a means of determining shibboleths within each group. While these cluster determinants are informative in their own right, the associated distance measures might also be seen as providing a sense of the overall linguistic ‘coherence’ of each group. The mean within-distance measure—the average linguistic distance between members of a cluster for a given set of features—may reveal groups of speakers whose overall degree of internal differentiation remains quite high. In cases such as this, additional documentation may help determine whether the cluster in question is indeed characterized by substantial internal variation, or whether the addition of more speakers with similar sociodemographic backgrounds would result in the emergence of several more linguistically coherent subgroups. By the same token, clusters with low average between-distance scores (i.e., where speakers share many of their features with members of other clusters) and low average within-distance scores (i.e., where speakers are less than uniform in the features they have in common with other members of their cluster) might benefit from further scrutiny to determine whether or not they can be reasonably attributed other groups, instead. In both cases, dialectometric measures such as these can be used productively to guide ongoing documentation, bringing attention to groups that demonstrate particularly rich variation or that pose possible exceptions to existing models of variation.

In general, the above dialectometric methods bring to light a wealth of linguistic variation

in the Saskatchewan Valley in each aspect of linguistic organization considered here, and serve a valuable role in determining the association of such variation with particular groups of speakers in the region. Importantly, such groupings are not the result of dialectometric models having been presented with these divisions in advance. Their coincidence with what has been reported in the literature on linguistic variation in Mennonite communities elsewhere and what is known about the sociodemographic history of the Saskatchewan Valley area only adds further plausibility to these inductively derived clusters. Moreover, the ability of these dialectometric methods to aggregate over large numbers of linguistic variables without favouring any one subset in advance is particularly useful in the present case, where relatively little is known about variation in the region.

These same methods also draw attention to features of variation in these communities that have received less attention in the literature on Mennonite Plautdietsch, to the point of being all but unattested. Here, variants that have commonly been reported to have categorical, shibboleth-like distributions separating particular groups of speakers are found to demonstrate considerable intra-speaker variation, as with the variation in *-e(n)* endings among speakers of both the central and transitional clusters. Associations between features such as this and particular clusters of speakers can be assessed with dialectometric methods even in the face of recurrent intra-speaker variation, where an absolute, isogloss-like patterning of variants between speakers may not be present. In these cases, features with both exclusive and non-exclusive distributions across clusters still emerge as part of the profiles of variation associated with particular groups of speakers.

Given the apparent exuberance of variation in the Saskatchewan Valley, the results of these analyses provide valuable insight into the overall patterning of variation in the region. The task of situating these results more precisely in the context of the sociohistorical and disciplinary questions that were set out in preceding chapters, as well as bringing attention to areas in which other analytical methods may contribute to interpreting the distribution of variation seen here, is therefore taken up in the concluding discussion of the following chapter.

## 6 Conclusion

The preceding chapter presented a documentation-based, quantitative analysis of linguistic variation as attested in the *Fibel* Corpus, drawing on dialectometric methods to identify and associate recurring patterns of variant usage with particular regions and groups of speakers. The following sections consider this description of variation in the Saskatchewan Valley in broader terms, attempting to assess its relevance to the methodological and analytical goals set out in earlier chapters (§6.1), its relationship to prior reports concerning variation in Mennonite Plautdietsch speech communities and to Russian Mennonite history more generally, and its implications for recent theoretical models of variation in communities with similar patterns of social organization (§6.2). As the conclusion to this chapter argues, while the concentration of this study on the synchronic linguistic practices of a smaller speech community such as this may have initially appeared unduly narrow in scope, this perspective belies the potential contributions that such communities have to make to current linguistic research in both its theoretical and methodological dimensions (§6.3).

### 6.1 *Methodological aims*

As the previous chapter has made apparent, linguistic variation in Saskatchewan Valley Mennonite Plautdietsch is complex, but not incoherent. The dialectometric methods employed in Section 5.2 bring attention to considerable structure in the variation attested in the *Fibel* Corpus in its linguistic and geographic (and, by extension, social) dimensions. On the basis of the aggregate patterning of such variation across several areas of linguistic organization, these methods have been able to suggest linguistically and sociodemographically plausible groups of speakers that share conventional patterns of variant usage, without privileging any one kind of variation in advance as being necessarily more determinant or of greater potential interest than the others. Associated analytical techniques have, in turn, added further detail to the profiles of each group that inductively emerge under these methods, highlighting the distinctive and common linguistic characteristics of each. The result of this analysis is an initial description of synchronic linguistic variation in the Saskatchewan Valley that establishes an outline of both the range of linguistic variation attested in these communities and its distribution across the local speaker population.

At this point, it is reasonable to consider the adequacy of this description in addressing the questions that were raised in Chapter 4 and explored in greater detail in Chapter 5. The first of these questions concerned the descriptive and documentary goals of this study, aiming to determine the range of variation found in present-day Saskatchewan Valley Plautdietsch and to find suitable forms of representation for it that might elucidate the linguistic situation and inform further documentation. The dialectometric methods adopted in the preceding chapter have sought to accomplish this, considering from the outset the full range of variation attested in the corpus rather than limiting its analytical perspective prematurely by focusing on a smaller set of features that may have been hypothesized to be of particular discriminatory or theoretical value. Indeed, the ability of dialectometric methods to aggregate over large numbers of variables and to visualize the distribution of individual features was particularly valuable in this respect, allowing for descriptive attention to be given to the overall patterning of variation without occluding the individual variables that contribute to such larger trends. As Section 5.3 noted, these methods not only served to establish the range and distribution of variation across the region, but also brought attention to areas in which additional documentation may be particularly useful (e.g., with clusters of speakers demonstrating considerable internal variation). In this respect, dialectometric methods appear to have generally satisfied the aims of this first question, providing one means of approaching attested variation in such a way as to offer direction to continued documentary efforts.

The second of these guiding questions focused on the relationship between linguistic variation and the demographic and historical characteristics of local communities and their members. It must be acknowledged that the dialectometric methods adopted in the preceding chapter have been somewhat more limited in addressing these points. While the visualizations and additional analytical techniques provided by current dialectometric tools present a clear view of the geographical dimensions of this question, and provide no small number of methods with which to explore the relationship between physical geography and observed patterns in linguistic variation, they are considerably more restricted in their ability to relate these observations directly to demographic and historical features that may also be of interest. The focus of dialectometric methods on modelling the properties of sites, rather than of speakers, contributes in part to these difficulties. Adapting these methods to bring additional non-linguistic, non-

geographical factors to bear on the overall patterning of variation remains an open problem for present tools in dialectometry.

To address these shortcomings, the preceding sections attempted to relate the geographical clusters of speakers identified by dialectometric means with the demographic and historical features associated with members of each such group. Although not directly supported by the functions provided in current dialectometric tools, nothing in the present approach prevented further correlations between the observed clusters and other sociodemographic or historical features from being explored separately. Thus, it was noted that the ‘central’ and ‘peripheral’ clusters of speakers corresponded not only with certain regions in the Saskatchewan Valley, but also with particular denominations and waves of migration that brought Mennonite settlers to those areas. In a similar way, Table 29 and subsequent discussion related the clusters of speakers arrived at under aggregate dialectometric analysis with other, non-geographical and non-linguistic features of interest. Although these approaches were generally informal and exploratory, they nevertheless provided another means of incorporating other potential predictors of variation into analysis that would otherwise have received less attention with dialectometric methods alone.

Other methods might also be adopted in future studies to complement the dialectometric approach taken here to addressing the second of these questions. Although the relatively large number of variables ( $n = 105$ ) and comparatively small number of respondents ( $m = 42$ ) may pose problems for some forms of quantitative analysis that require larger amounts of data to estimate the values of each included model parameter (e.g., linear regression), one might first attempt to reduce the dimensionality of the present data with non-parametric ensemble classification methods such as random forests (Strobl et al. 2009). These methods apply traditional classification and regression tree (CART) methods to a large number of random subsets of data and predictors to produce a ‘forest’ of classification trees, each of which establishes the relative importance of individual variables in predicting a classificatory outcome of interest (e.g., speaker cluster). By comparing the relevance of each variable to predicting the outcome across all of the available classification trees, random forest-based methods are able to produce a measure of the relative importance of each predictor to classification. Applying these methods to the data from the Saskatchewan Valley would provide an indication of the variables

most relevant to arriving at the observed clusters of speakers, and thus offer a more informed way of winnowing out less predictive variables before pursuing other forms of analysis.

In addition to ensemble classification techniques, several other families of quantitative methods may also complement the dialectometric approach taken in the preceding chapter. In particular, both Multiple Correspondence Analysis (MCA) and Generalized Additive Modelling (GAM) show promise in bringing both linguistic and non-linguistic features to bear on the analysis of variation. MCA is an exploratory technique that extends Principal Components Analysis to large sets of categorical variables, reducing their dimensionality to two or three-dimensional Euclidean spaces in which the relationships between respondents and variables can be more readily visualized and interpreted (Husson, Josse & Pagès 2010, Tummers, Speelman & Geeraerts 2012). Although MCA does not appear to have been applied to sociolinguistic or dialectological problems of this kind in previous linguistic research (or, indeed, in much linguistic research at all, with the notable exceptions of Glynn 2009, Glynn 2010, and Schrauf 2013), its facility with large numbers of categorical predictors and the ease with which its results may be interpreted make it an attractive counterpart to the present methods. By comparison, recent applications of GAM to larger collections of dialectological data (e.g., Wieling, Nerbonne & Baayen 2011) demonstrate the effectiveness of these methods in exploring similar multivariate questions, combining the advantages of mixed-effects regression modelling with the explanatory contributions of geographical predictors. Preliminary applications of both GAM and MCA to data from the *Fibel* Corpus suggest that they are both technically feasible, although a discussion of their results falls outside of the scope of this study. In general, methods such as these might be expected to complement those provided by dialectometry, particularly in treating potential relationships between linguistic and sociohistorical factors more directly.

The third guiding question raised in the preceding chapters dealt with the possibility of identifying coherent, recurring patterns of variation inductively, without imposing such structure upon the data *a priori*. The clusters of speakers reported in Chapter 5 satisfy both of these requirements, presenting patterns of variant selection that are attested across multiple speakers in the Saskatchewan Valley and which were arrived at in a bottom-up fashion through agglomerative clustering. While this framing of research interests focused on finding structure specifically in the linguistic form that such variation assumed, it bears noting that the clusters of

speakers who employ these patterns of variant selection also show notable geographical (and, to some extent, demographic and historical) coherence, as discussed above. These factors lend additional plausibility to such formally derived clusters, suggesting that they are not merely chance associations between otherwise unrelated speakers, but that they have some basis in the social and historical reality of the region.

Finally, the fourth question attended to the means and outcomes of linguistic research undertaken in contexts similar to the Saskatchewan Valley, where processes of language shift and loss are already well underway and relatively few resources on the traditional linguistic practices of the community are available either to academic research or to local language initiatives. While the methods by which this study has attempted to be responsive to the interests of multiple stakeholders have already been discussed at several points (cf. §1.2.4, §4.2, §4.2.2.2, §4.2.3.1), it is perhaps appropriate to note that both the more tangible products of research (i.e., the *Fibel* itself) and the less tangible results of the preceding quantitative analysis may be relevant to non-academic partners. While the *Fibel* has been explicitly designed to serve as a resource for those interested in the language, the description presented here of the range and character of variation in Saskatchewan Valley Mennonite Plautdietsch may also be relevant to local language planning (e.g., in determining how further resources and programs should be developed in order to adequately reflect the diversity of local linguistic practices). Continued discussion with community stakeholders will be needed to determine how these results might best be shared, such that they are maximally accessible to all project partners.

This approach to assessing the form and distribution of linguistic variation in the Saskatchewan Valley thus addressed all four sets of criteria that were identified as priorities for analysis in preceding chapters. Moreover, it should be noted that it was possible for this analysis to be conducted directly on the basis of documentary data, with the results of analysis traceable to individual responses in the *Fibel* Corpus. With all stages of analysis—from the extraction of concordance lines from the corpus to their eventual transfer into dialectometric tools—having been recorded in separate scripts, the results of this study are amenable to further expansion (e.g., to comparative data from other communities, the recoding or reanalysis of the present corpus data, or novel sources of information on local linguistic practices derived from other sources) and to independent replication. As was noted in Section 1.2.4, this kind of analytical reuse and

enrichment of primary data is in keeping with recommendations in documentary linguistics, which commonly stress the importance of linguistic descriptions being based upon and contributing to the interpretation of permanent documentary resources. The results of this study suggest that such documentation-based, reproduceable approaches to linguistic description are indeed viable, even with underdocumented languages such as Plautdietsch and in cases where relatively little is known in advance about the extent or distribution of significant features of the linguistic practices of interest.

While perhaps successful in these respects, the present study remains only an initial description of linguistic variation in the Saskatchewan Valley, and leaves several issues for continued research in this area. From a methodological perspective, it is apparent from this analysis that the translation task-based approach adopted here, while providing systematic and tractable representation of a relatively large number of variables from multiple areas of linguistic organization across a range of speakers, is not without significant limitations. As preceding sections have noted, such methods are clearly not viable in all Mennonite Plautdietsch speech communities, as English bilingualism is not a constant feature across the Russian Mennonite diaspora (although see Kaufmann 2005 for discussion of the challenges of employing a similar survey instrument in translation across several diasporic Russian Mennonite communities). They also show notable limitations with syntactic and prosodic data, where the nature of the translation task appears to result less consistently in reliable data (cf. §4.2.1.5). For these reasons, the translation-based approach adopted here would benefit substantially from the counterbalance afforded by semi-controlled linguistic tasks (e.g., the narration of a wordless picture book or video, which provides a common task and stimulus set for comparison across speakers and communities while allowing for considerable individual linguistic creativity; cf. Bower 2008: 82–83, Lüpke 2009) and uncontrolled linguistic tasks (e.g., unscripted interviews and conversations) in the same communities. Information from methods such as these would greatly enrich the contents of the *Fibel*, allowing for a clearer assessment of the degree to which the linguistic task itself may have influenced speakers' responses, and contribute to the development of other language resources that may be of interest to both academic and non-academic research partners.

A broader methodological approach to understanding linguistic variation in Mennonite



Plautdietsch might also allow these investigations to be extended to Plautdietsch speech communities outside of the Saskatchewan Valley. Further comparison with other diasporic communities would contribute significantly to the documentation of these other varieties, many of which remain entirely undocumented. At the same time, such research would also afford a clearer perspective on the particular linguistic features that set these communities apart from one another. Even the limited comparison made in the present study between Mennonite and Catholic Plautdietsch communities in Saskatchewan makes the benefits of such comparison apparent, providing greater insight into both the shared and distinctive linguistic characteristics of each group.

Systematic comparisons with other Plautdietsch speech communities may also present an opportunity to extend the focus of this research from patterns of synchronic variation to larger, diachronic processes of language change that may be reflected in the linguistic conventions of each group. In the case of Mennonite Plautdietsch speech communities with relatively recent historical ties to the Saskatchewan Valley (e.g., those located in southeastern and northern Saskatchewan, northern Alberta, and British Columbia, or, more distantly, in northern Mexico, Belize, and Bolivia), such research could contribute to the identification of instances of apparent linguistic change in progress. Similarly, comparisons with non-Mennonite Plautdietsch communities, although much fewer in number, presents a rare opportunity to gain comparative perspective on the range of linguistic features that distinguish each group, providing information that may shed light on earlier stages in the development of both Mennonite and non-Mennonite varieties of Plautdietsch. As the following sections address, such investigations have much to contribute to our understanding of how linguistic variation in Mennonite Plautdietsch speech communities patterns, and how this relates to the history and social structure of these groups in general.

## 6.2 *Variation in Mennonite Plautdietsch revisited*

Chapters 2 and 3 gave attention to the history of the Russian Mennonites and to research into their linguistic practices. While this discussion initially served to situate the Mennonite communities in the Saskatchewan Valley in clearer historical and linguistic context, the results of the analysis that followed this introduction also have some bearing on questions in Mennonite

history and claims that have been made concerning the linguistic practices of groups throughout the Russian Mennonite diaspora. This section pursues several such points of connection further, considering in particular the relationship between the observations of the preceding analysis and existing models of linguistic variation in Mennonite Plautdietsch, conventional methods for its investigation, and hypotheses concerning its emergence and maintenance.

Among the most prominent features of variation identified in the Saskatchewan Valley was a deep bifurcation in contributors' patterns of variant usage. In the preceding analysis, the main branches of this division were provisionally assigned the labels 'peripheral' and 'central', referring to the areas of the Saskatchewan Valley in which each associated group of speakers was centred. Further investigation of these groups in Section 5.3 suggested additional, meaningful correlations between these patterns of variation and the migrational and denominational histories of associated speakers and their families. This was not entirely unexpected, given the history of these groups and the findings of previous research in historically related communities. In the studies reviewed in Chapter 3, such correlations have generally been attributed to one or another division assumed to be of primary explanatory value. Thus, such studies often identify period of emigration (*Kanadier* vs. *Russländer*), or colony of origin (Chortitza vs. Molochnaya), or dialect region (*Nehrung* vs. *Werder*), or denominational affiliation ('conservative' vs. 'progressive', or, historically, Flemish vs. Frisian) as the deciding feature in explaining such persistent binary divisions between Mennonite Plautdietsch varieties. The apparent recurrence of such a division in the Saskatchewan Valley presents an opportunity to consider how such labels have typically been applied to linguistic differences between diasporic Russian Mennonite groups. With such labels having been the dominant mode of discussing variation in Russian Mennonite communities to date, comparison among their use and the results of the analysis here might serve either to relate the patterns of variant usage observed in the Saskatchewan Valley to patterns reported in other communities—or, alternatively, to identify areas in which such reports differ from the conclusions drawn here.

As prominent as this linguistic division between 'peripheral' and 'central' speakers in the Saskatchewan Valley may be, it is clear from the preceding analysis that geography is not the sole possible explanation for this difference. Indeed, as noted in Section 5.3, it is difficult to square these labels directly with any one set of the existing analytical categories mentioned

above for several reasons. First, in the present case, geography, migration, and denominational affiliation are all highly correlated. Speakers of the ‘central’ cluster tend to be of Old Colony or Bergthaler denominational background whose ancestors emigrated from related settlements in Manitoba and Ukraine in the same time period, while speakers in the ‘peripheral’ cluster tend to be the descendants of later Mennonite immigrants from other denominations. This is common elsewhere in the Russian Mennonite diaspora, as well, with many earlier migrations involving groups of families, typically of the same denomination and often with significant kinship ties between them, relocating to establish new settlements together. In the absence of further evidence, it would be difficult to conclude that any one of these factors is primarily responsible for the observed linguistic differences—nor is it perhaps realistic to expect that a single-factor explanation is able to account for all such variation, given the tight bundling of these features throughout Russian Mennonite history.

Nevertheless, such labels are common in studies of variation in Russian Mennonite speech communities, whether applied monofactorially to account for linguistic differences between groups or multifactorially to identify relevant facets of differentiation between them. One might question the assumed independence of these labels in the latter cases, as well, given the cross-cutting and interrelated nature of many of these factors. Dyck (1964: 57ff.), for instance, presents extensive coverage of individual features reported to differ between speakers of ‘Chortitza’ and ‘Molochnaya’ background. Although the intended meaning of these labels is clear, their application in the absence of other correlated features (e.g., in a Canadian context, the period during which speakers from these colonies emigrated from Ukraine) is potentially problematic. As the results of this study reinforce, the relative degree of distance between ‘Chortitza’ and ‘Molochnaya’ ways of speaking does not appear to be constant over time periods or waves of emigration; there is reason to believe that linguistic differences between both colonies diminished notably in the period between the first Russian Mennonite migrations to North America and the end of the Russian Revolution. It cannot be assumed that the sets of linguistic conventions to which such labels refer are either diachronically stable or independently meaningful without further conditioning sociohistorical information. When applied in isolation to refer to groups of speakers, their utility as linguistic labels would appear at best limited, and at worst contribute to an empirically unwarranted reduction of the range of factors shown to be

relevant to such variation.

Thus, when used to characterize the linguistic conventions of entire groups of speakers without further supporting detail, such labels may inadvertently suggest monolithic patterns of variation that are not borne out in actual observation. Applying the labels ‘Molochnaya’ or ‘Chortitza’ to refer to the patterns of speech associated with these colonies, as Dyck (1964: 57) does, implies a degree of uniformity in the distribution of linguistic features on either side of this proposed dividing line; speakers of one colonial background will assumedly use feature A, while speakers of the other will use feature B. In the Saskatchewan Valley, however, there is little evidence that the ‘central’ and ‘peripheral’ clusters are internally homogeneous in a similar way. On the contrary, intra-group variation is robustly attested in each of these clusters and is often not easily attributed to the same set of factors that are claimed to set these speakers apart from the members of other clusters (cf. §5.3 on the patterning of siblings within their dialect groups). Although there is both linguistic and onomastic evidence of a long-standing division between two broad groups within the Russisan Mennonite diaspora (cf. Schapansky 2006), these ‘two solitudes’ in the same religious community, wherever the contemporary placement of the major societal dividing line between them may fall—whether denominationally between Flemish or Frisian and conservative and progressive, or geographically between *Nehrung* or *Werder*, or colonially between Chortitza or Molochnaya, or emigrationaly between *Kanadier* or *Russländer*—has clearly not evinced a perfect separation, nor prevented the maintenance of significant linguistic heterogeneity within each group. A narrow concentration on the historiographical labels assigned to such divisions risks diverting attention from both considerable intra-group variation and significant instances of inter-group contact that are of no lesser relevance to the patterns of language usage observed today.

The application of binary labels such as these has also contributed at times to less attention being paid to speakers who are not as easily assigned to either side of the posited division. Characterizing variation in diasporic Mennonite communities as a matter of ‘Chortitza vs. Molochnaya’ or ‘*Russländer* vs. *Kanadier*’ may leave speakers who have associations with both categories altogether out of frame. Yet, these individuals are often no less a part of the local speech community than others whose linguistic conventions fall more neatly into established categories. Leaving aside the significant differences observed between Mennonite and Catholic

varieties of Plautdietsch, the presence of so-called ‘transitional’ speakers among the Mennonite community in the Saskatchewan Valley alone calls into question the adequacy of a binary treatment of such variation. Under these kinds of analysis, groups such as these ‘transitional’ speakers are rarely treated as having conventional patterns of variation of their own, and are more often divided between one or the other established categories or given no attention at all.

Both of these issues—the treatment of local majorities as more or less linguistically and demographically homogeneous and the lessened attention given to speakers whose histories or patterns of use place them outside of these dominant groups—are recurring ones in studies of variation across the Russian Mennonite diaspora, and contribute in due part to the competition between the analyses summarized in Chapter 3. In the present study, at least, methods of analysis based on permanent documentation (rather than introspection or first-hand reports, as in Dyck 1964) and forms of representation in which variation outside of prominent dichotomies are not excluded (e.g., through methods of visualization that present a sense of variation within groups) are clearly feasible, and may present one means of approaching the description of variation that avoids some the difficulties associated with ‘all-or-nothing’ treatments of heterogeneous linguistic conventions.

Attempts to come to terms with linguistic variation in Mennonite communities through finer-grained classification have struggled with the considerable range of individual differences among speakers. Mitzka (1930: 22) gives an early summary of such efforts, and argues on this basis that any dialect differentiation of real significance is to be found between the varieties associated with the Chortitza and Molochnaya colonies:

*Die Versuche, das Mennonitenplatt in eine längere Reihe von Mundarten aufzuteilen, treffen leicht bloße Einzelheiten, nicht durchgehende Erscheinungen. [...] Wirklich durchgehende Unterschiede zeigen die Mundarten der Alt- und der Neu-Kolonie.*

Attempts to divide Mennonite Plautdietsch into a longer series of dialects arrive readily only at mere minutiae, rather than pervasive phenomena. [...] Truly pervasive differences are shown by the dialects of the Old and the New Colony.

Without calling into question the accuracy of this statement, it is telling that Mitzka’s characterization of the linguistic situation does not rule out the existence of other, assumedly less linguistically significant differences between Mennonite speech communities. While focusing

here on the latter, ‘pervasive’ differences, Mitzka tacitly acknowledges a range of other variation that he suggests to be of lesser classificatory relevance. Given the importance placed in this research programme on structural phonological and morphological differences of a Neo-Grammarians kind, it is not surprising that other forms of variation might at times be disregarded as “mere minutiae,” presenting at best a disappointing day’s catch on the sea of linguistic diversity.

Yet, in the present study, applying the same degree of attention to such seemingly minor features as to the structural differences prized by Mitzka reveals no less coherent linguistic boundaries in these features between speaker groups, even when individual variants are less distinctive in their distribution. While perhaps less striking than the major phonological shifts or recurring morphological differences to which Mitzka primarily attended, and even when shared to some degree between groups of speakers, taken together, such ‘minor’ features nevertheless silhouette larger and more coherent conventions of variation than Mitzka’s statement attributes to them. In the Saskatchewan Valley, this was most prominently the case with speakers in the so-called ‘transitional’ group. While these transitional speakers shared many variants with several other clusters of speakers, thus decreasing the likelihood of their treatment as a distinct group in their own right under the present methods, the combined information provided by other, seemingly inconsequential features nevertheless presented evidence of an overall profile of conventional variation that differed substantially from other populations. With these speakers, it was the constellation of a host of non-categorical, non-dichotomous features that formed the boundary with other constituent groups of the larger speech community. The combined occurrence of such apparently minor and seemingly unrelated features, then, displayed much the same hallmarks of conventionality as the sharp morphophonological dividing lines that are typically favoured in analysis, with the mosaic patterning of many smaller features—all in characteristic alignment and with no one feature being necessarily distinctive on its own—together defining the larger profile of a linguistic variety.

Assessments such as Mitzka’s arguably reflect a subtle bias in linguistic analysis, both then and now, towards certain forms of structural variation, favouring variables with particularly uneven distributions across sociodemographic categories while disfavouring variables whose individual patterns of use are seen as less discriminative in arriving at the categories of interest.

Indeed, even the cluster determinant metrics applied in Section 5.2 sought to minimize presence of shared variables in its list of discriminant features, instead attempting to arrive at all and only those variables whose distributions were markedly different between groups. While such shibboleths are no doubt important linguistic features, other approaches to understanding variation, whether through the visualizations provided by dialectometric methods or the profiles of variation afforded by Multiple Correspondence Analysis, allow other, perhaps less obviously interrelated factors to remain part of the larger analytical picture. From the standpoint of the descriptive aims of this study, it is reasonable to prefer alternatives to jettisoning such information, both for the overall empirical adequacy of the resulting description as well as for the further detail it provides concerning the overall linguistic situation in the region. Such features are arguably no less part of the fabric of convention in local speech communities than the more structurally prominent and sociodemographically skewed variables that are commonly at the centre of linguistic analysis.

Other observations made in the present analysis are also relevant to outstanding claims in the literature on Mennonite Plautdietsch. The intra-speaker variation noted in certain linguistic features, such as the *-e(n)* endings reviewed in Section 5.2.3, is particularly noteworthy, as these variables have received considerable attention as indicators of major dialect boundaries in preceding studies. The attestation of such variation among multiple groups of speakers in the Saskatchewan Valley calls into question the treatment of these variables as inter-group markers only and merits further attention. As well, the considerable linguistic differences between siblings observed in Section 5.3 suggests that the model of family-based transmission proposed by Moelleken (1967) cannot be the entire story of how variation in Mennonite Plautdietsch speech communities is maintained across generations. While Moelleken bases his conclusion that “the family, not the village, is the smallest unit capable of retaining speech realizations” (i.e., coherent sets of linguistic conventions; Moelleken 1967: 251) on differences in the phonological inventories of *Russländer* Mennonite immigrants to Canada from the same Ukrainian villages, no less significant variation between individuals belonging to the same family is observed in the Saskatchewan Valley settlements. If siblings here bear no greater linguistic resemblance to one another than they do to other members of their same dialect group, then other factors must be at play, as discussed in the following section.

### 6.3 *Linguistic variation and social structure*

When considering the results of the preceding chapter as a whole, what is perhaps most striking is the extent of linguistic variation represented in a relatively small speaker population. Indeed, at first blush, the persistence of such extensive variation over several generations of Mennonite settlement in Canada appears remarkable, especially given the close-knit nature of these communities. Although there are some signs of linguistic accommodation among speakers separated from others of the same background to locally dominant vernacular norms, both in the Saskatchewan Valley and elsewhere in the Russian Mennonite diaspora, apart from these cases, there is little evidence of significant convergence to a single linguistic norm over multiple generations of local settlement. This is contrary to the expectations of models of dialect levelling and new dialect formation such as that of Trudgill (1999), where the emergence of a stable, levelled dialect would be normally anticipated in the third generation of speakers when inter-speaker and intra-speaker variation found in the preceding generation converges. Instead, considerable variation at the level of the individual remains well attested in Canadian Mennonite communities several generations on, to the extent that pairs of siblings in the present sample are no more linguistically similar to one another than they are to any other speaker of the same broad dialect background (cf. §5.3)—with little influence apparently exerted by factors of age, gender, or place of birth on individuals' linguistic practices within their respective dialect group.

The persistence of significant, individual-level variation across multiple generations of Mennonite settlement, showing little sign of diminishment or convergence to a single norm and no marked differences between ages and genders, bears a strong resemblance to similar reports of non-convergent, personally patterned variation presented in Dorian (2010). In the case of East Sutherland Gaelic, on which Dorian (2010) concentrates, but also in similar reports from other speech communities internationally (cf. Dorian 2010: 271–313), such persistent, idiosyncratic variation is argued not to be solely the result of linguistic obsolescence or of an incomplete process of dialect formation, but rather a consequence of a particular form of social organization that lessens normative pressure towards linguistic accommodation, allowing significant inter-speaker and intra-speaker variation to remain socially neutral. According to Dorian (2010: 286), several factors are likely contributors to the development and persistence of such variation:



- dialect mixture via population mixing;
- geographical isolation and/or enclavement within a larger allophone population;
- minority status for the language in question, with either no relationship or distant relationship to the dominant language of the country or region;
- absence of community-external language norms and exclusion of the minority language from written use among its speakers;
- absence of social stratification related to socioeconomic differentiation;
- absence of social evaluation of variants vis-à-vis one another;
- a homogeneous small-community social structure characterized by dense face-to-face interaction and multiplex social roles;
- absence of linguistic accommodation;
- obsolescence (with declining use an exacerbating, if not originating, factor)

Notably, as Dorian (2010: 286–287) points out, these same features are almost exactly the opposite of the settings in which the majority of linguists and linguistic anthropologists are socialized, where social stratification is commonplace, the dominant languages are written and used in education, relationships are rarely multiplex, and the realizations of particular linguistic variables are often associated with membership in a particular class or ethnic group. Settings in which high degrees of linguistic variation are essentially socially neutral, as Dorian (2010) argues is the case for East Sutherland Gaelic, are thus contrary to the expectations that many researchers implicitly bring to the task of linguistic description. Rather than being solely a consequence of language shift or incomplete convergence, Dorian (2010: 287) argues that abundant variation with minimal social weighting is itself an entirely “reasonable outcome of a particular set of social, historical, and linguistic features” that favour the emergence of such norms—features which, she continues, “should make us alert to the possibility of social and linguistic conditions with consequences we have not fully appreciated previously” (*ibid.*).

The social and linguistic factors that Dorian identifies as favouring the emergence of persistent idiosyncratic variation apply without exception to the Plautdietsch-speaking population of the Saskatchewan Valley, as well. In these enclaved minority speech communities, significant population mixing is attested, both through multiple historical waves of Mennonite migration and through the subsequent dynamics of agricultural expansion and the gradual weakening of earlier denominational and emigrational divisions. Moreover, Plautdietsch in the Saskatchewan Valley is rarely used in written communication, is significantly affected by

ongoing language shift, and bears only a distant relationship to the dominant English, with no normative linguistic pressure exerted by this language (or, in most cases, from those distinct varieties of Standard German maintained within the Mennonite community; cf. Cox 2013) on local vernacular practices.

In societal terms, then, the Mennonite communities in the region are well aligned with the conditions proposed by Dorian to lead to significant levels of socially neutral variation. Until the occupational diversification in Canadian Mennonite society that followed the Great Depression and Second World War, Mennonite communities in the Saskatchewan Valley exhibited little socioeconomic differentiation, with virtually all members of the community being employed in similar forms of small-scale agriculture (cf. §§2.4–2.6). Social stratification in other important aspects of Mennonite life was arguably further limited by the relatively ahierarchical nature of Mennonite denominations, in which even elected ministers and bishops (*Ältester*) continued to manage their own farms as any other member of the community. Moreover, village-based settlements established by particular denominations or groups of families with close social ties typically connected their inhabitants not only as neighbours, but often also as members of the same church and, more often than not, through vast and intimate networks of kinship and face-to-face visiting that bound together Mennonite families throughout the region. Outside of certain features associated in local communities with the division between *Kanadier* and *Russländer* groups, there is little evidence of marked social evaluation of linguistic variants in either group. While this dividing line remained prominent and subject to considerable commentary in the local community, little normative attention appears to have been given to variation *within* either group, and linguistic accommodation appears to have been limited only to occasional cases of contact between these historically separate communities.

Variation in Saskatchewan Valley Mennonite Plautdietsch, then, may have further relevance as a point of comparison for the claims made by Dorian (2010) concerning the relationship between linguistic variation and social structure. As that study repeatedly notes, societies in which the conditions for diminished sociolinguistic evaluation and heightened levels of persistent idiosyncratic variation are met—those with relatively small populations, dense and multiplex networks of interrelation and interaction, general ethnic and socioeconomic homogeneity, and weak extra-community linguistic norms—are now increasingly uncommon,

representing exceptions to the forms of social organization that have become dominant in much of the world. Other recent studies of variation in smaller speech communities with forms of social organization differing markedly from much of Western society (e.g., Stanford 2012) similarly suggest that conclusions of synchronic and diachronic uniformity in the patterning of sociolinguistic variation must ultimately be evaluated in the context of the particular societies in which such variation is embedded, rather than attempting to extrapolate universal tendencies from the relatively limited number of societal configurations that are commonly examined. Smaller, densely interconnected, and minimally stratified speech communities such as those in the Saskatchewan Valley may thus present a rare contemporary test case for current theories of linguistic variation and social structure, contributing to a more adequate typology of language use in relation to forms of social organization.

#### **6.4**    *Conclusion*

Engagement with smaller-language communities such as those in the Saskatchewan Valley has much to contribute to contemporary linguistics—far more than what might be expected, given the attendant challenges that the progression of language endangerment and the lack of prior documentation may pose (cf. Whalen 2004). In the present context, issues such as these belie the potential contributions that such communities might make to current linguistic and sociolinguistic typology, as noted in the preceding section, and encourage further consideration of areas in which contemporary linguistic practices themselves may be in need of refinement. As an example, the lack of previous documentation and expressed interest in resources for language education in the Saskatchewan Valley encouraged a collaborative model of documentation in this study, ultimately resulting in a resource relevant both to academic linguistic research into synchronic variation and to local language promotion and revitalization efforts. This approach also served to bring attention to areas in which current linguistic tools and common workflows fall short of supporting the development of multi-purpose documentation in contexts such as these, pointing out instances where custom software development was required to accomplish even relatively straightforward tasks (e.g., conversion between common data formats, or the batch annotation of information distributed across multiple records in a collection). Such challenges are far from marginal and are certainly not limited to the

Saskatchewan Valley alone. The engagement of linguistic research in contexts such as this thus presents opportunities not only to seek to advance the present state of linguistic knowledge, but also to reflect critically on the applicability of current disciplinary practices to increasingly common situations of linguistic endangerment and underdocumentation.

As was suggested in the introduction to this study, the history of the Saskatchewan Valley can be seen to present an almost synecdochic reflection of the Russian Mennonite story as a whole, a place where the manifold strands of centuries of emigration and exile have met and again become intertwined. In this light, linguistic research offers another perspective on the complex history of these communities, complementing the observations of other forms of scholarship with an understanding of how the dynamics of division and contact are made manifest in the linguistic practices of the region. It would be difficult, if not impossible, to separate such observations about the linguistic conventions of these communities from their social and historical embedding. Nor, arguably, should it be the aim of linguistic analysis to set its contributions apart from the interests of related disciplines and local communities with which it shares a stake in understanding the present situation. In the collaborative, community-partnered methods adopted here, this study suggests that another way forward is possible—one in which linguistic analysis contributes not only to the advancement of the discipline, but also to a common appreciation of the importance of such communities' linguistic practices and to local efforts to see them remain vibrant in their full diversity into the future.

## References

- ALLEN, WILL, JOAN BEAL, KAREN CORRIGAN, WARREN MAGUIRE, and HERMANN MOISL. 2007. A linguistic 'time capsule': the Newcastle Electronic Corpus of Tyneside English. In BEAL, JOAN C., KAREN P. CORRIGAN, and HERMANN L. MOISL (eds.), *Creating and digitizing language corpora. Volume 2: diachronic databases*, 16–48. New York / Basingstoke, Hampshire: Palgrave-Macmillan.
- ANDERSON, ALAN B. 2005. *German settlements in Saskatchewan*. Saskatoon, Saskatchewan: Saskatchewan German Council.
- BAERG, MARJORIE LANE. 1960. *Gnadenau Low German: a dialect of Marion County, Kansas*. Unpublished doctoral dissertation, University of Chicago.
- BAINTON, ROLAND H. 1952. *The Reformation of the sixteenth century*. Boston: Beacon Press.
- BARBIERS, SJEF, HANS BENNIS, MAGDA DEVOS, GUNTHER DE VOGELAER, and MARGREET VAN DER HAM (eds.). 2005. *Syntactic atlas of Netherlandic dialects. Volume 1*. Amsterdam: Amsterdam University Press.
- BARBIERS, SJEF, LEONIE CORNIPS, and JAN PIETER KUNST. 2007. The Syntactic Atlas of the Dutch Dialects (SAND): a corpus of elicited speech and text as an online dynamic atlas. In BEAL, JOAN C., KAREN P. CORRIGAN, and HERMANN L. MOISL (eds.), *Creating and digitizing language corpora. Volume 1: synchronic databases*, 54–90. New York / Basingstoke, Hampshire: Palgrave-Macmillan.
- BARBOUR, STEPHEN and PATRICK STEVENSON. 1998. *Variation im Deutschen: Soziolinguistische Perspektiven* [Variation in German: sociolinguistic perspectives]. Berlin: Walter de Gruyter.
- BAUER, LAURIE. 2002. Inferring variation and change from public corpora. In CHAMBERS, J. K., PETER TRUDGILL, and NATALIE SCHILLING-ESTES (eds.), *The handbook of language variation and change*, 97–114. Malden, Massachusetts: Blackwell.
- BAUMANN, WILHELM. 1854. Mundarten der deutschen Ansiedler an den Ufern des Molotschnaja-Flusses im Taurischen Gouvernement im südlichen Rußland [Dialects of the German settlers on the banks of the Molochnaya River in the Taurida governorate in southern Russia]. In FIRMENICH, JOHANN MATTHIAS (ed.), *Germaniens Völkerstimmen III*, 434–445. Berlin: Schlesinger.

- BECH, GUNNAR. 1955. *Studien über das deutsche Verbum Infinitum* [Studies of the German infinitival verb]. 1. Band. Det Kongelige Danske Videnskabernes Selskab, Historisk-filologiske Meddelelser, bind 35, nr. 2. København: Ejnar Munksgaard.
- BECH, GUNNAR. 1957. *Studien über das deutsche Verbum Infinitum* [Studies of the German infinitival verb]. 2. Band. Det Kongelige Danske Videnskabernes Selskab, Historisk-filologiske Meddelelser, bind 35, nr. 2. København: Ejnar Munksgaard.
- BEREZ, ANDREA L. 2007. Software review of EUDICO Linguistic Annotator (ELAN). *Language Documentation and Conservation* 1(2), 283–289.
- BIBER, DOUGLAS. 1993. Representativeness in corpus design. *Literary and Linguistic Computing* 8(4). 243–257.
- BIRD, STEVEN and MARK LIBERMAN. 2001. A formal framework for linguistic annotation. *Speech Communication* 33(1–2). 23–60.
- BOAS, HANS CHRISTIAN. 2007. From the field to the web: implementing best-practice recommendations in documentary linguistics. *Language Resources and Evaluation* 40. 153–174.
- BOAS, HANS C. 2009. *The life and death of Texas German*. Durham, North Carolina: Duke University Press.
- BOWERN, CLAIRE. 2008. *Linguistic fieldwork: a practical guide*. New York: Palgrave Macmillan.
- BRANDT, CARSTEN. 1992. *Sprache und Sprachgebrauch der Mennoniten in Mexiko* [Language and language use of the Mennonites in Mexico]. Schriftenreihe der Kommission für ostdeutsche Volkskunde in der Deutschen Gesellschaft für Volkskunde e.v. 61. Marburg: N. G. Elwert Verlag.
- BREDNICH, ROLF WILHELM. 1977. *Mennonite folklife and folklore: a preliminary report*. Mercury Series Paper No. 22. Ottawa, Ontario: Canadian Centre for Folk Culture Studies, National Museum of Man.
- BREDNICH, ROLF WILHELM. 1981. *The Bible and the plough: the lives of a Hutterite minister and a Mennonite farmer*. Mercury Series Paper No. 37. Ottawa, Ontario: Canadian Centre for Folk Culture Studies, National Museum of Man.
- BRITAIN, DAVID. 2010. Conceptualizations of geographic space in linguistics. In LAMELI, ALFRED, ROLAND KEHREIN, and STEFAN RABANUS (eds.), *Language and space: an international*

- handbook of linguistic variation. Volume 2: language mapping*, 69–97. Berlin: Mouton de Gruyter.
- BRUGMAN, HENNIE and ALBERT RUSSEL. 2004. Annotating multi-media / multimodal resources with ELAN. In LINO, MARIA TERESA, MARIA FRANCISCA XAVIER, FÁTIMA FERREIRA, RUTE COSTA, RAQUEL SILVA (eds.), with CARLA PEREIRA, FILIPA CARVALHO, MILENE LOPES, MÓNICA CATARINO, and SÉRGIO BARROS, *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004), May 26–28, 2004, Lisbon, Portugal*, 2065–2068. Online: <http://www.lrec-conf.org/proceedings/lrec2004/pdf/480.pdf>
- BUCHHEIT, ROBERT H. 1978. *Mennonite 'Plautdietsch': a phonological and morphological description of a settlement dialect in York and Hamilton Counties, Nebraska*. Unpublished doctoral dissertation, University of Nebraska.
- BUCHHEIT, ROBERT H. 1982. Language maintenance and shift among Mennonites in south-central Kansas. *Society for German-American Studies Yearbook* 17. 111–121.
- BUCHHEIT, ROBERT H. 1988. Language shift in the concentrated Mennonite district of Kansas. *International Journal of the Sociology of Language* 69. 5–18.
- BYBEE, JOAN. 2006. *Frequency of use and the organization of language*. Oxford: Oxford University Press.
- CHAMBERS, J. K. and PETER TRUDGILL. 1998. *Dialectology*. 2nd edn. Cambridge: Cambridge University Press.
- CHILDS, BECKY, GERARD VAN HERK, and JENNIFER THORBURN. 2011. Safe harbour: ethics and accessibility in sociolinguistic corpus building. *Corpus Linguistics and Linguistic Theory* 7(1). 163–180.
- COX, CHRISTOPHER. 2008. *Verbal complementation in Mennonite Plautdietsch: a corpus-based, constructional approach*. Unpublished MSc thesis, University of Alberta.
- COX, CHRISTOPHER. 2010. Probabilistic tagging of minority language data: a case study using Qtag. In GRIES, STEFAN TH., STEFANIE WULFF, and MARK DAVIES (eds.), *Corpus linguistic applications: current studies, new directions*, 213–231. Amsterdam: Rodopi.
- COX, CHRISTOPHER. 2011a. Quantitative perspectives on syntactic variation: investigating verbal complementation in a corpus of Mennonite Plautdietsch. In KONOPKA, MAREK, JACQUELINE KUBCZAK, and ULRICH H. WASSNER (eds.), *Grammar and Corpora* 3, 549–555. Tübingen:

Gunter Narr Verlag.

- COX, CHRISTOPHER. 2011b. Corpus linguistics and language documentation: challenges for collaboration. In NEWMAN, JOHN, R. HARALD BAAYEN, and SALLY RICE (eds.), *Corpus-based studies in language use, language learning, and language documentation*, 239–264. Amsterdam: Rodopi.
- COX, CHRISTOPHER. 2013. The resilient word: linguistic preservation and innovation among the Old Colony Mennonites in Latin America. *Journal of Mennonite Studies* 31. 51–74.
- COX, CHRISTOPHER, JACOB M. DRIEDGER, and BENJAMIN V. TUCKER. 2013. Illustrations of the IPA: Mennonite Plautdietsch (Canadian Old Colony). *Journal of the International Phonetics Association* 43(2). 221–229.
- CRIPPEN, JAMES A. and LAURA C. ROBINSON. 2013. In defense of the lone wolf: collaboration in language documentation. *Language Documentation and Conservation* 7. 123–135.
- CROWDY, STEVE. 1993. Spoken corpus design. *Linguistic and Literary Computing* 8(4). 259–265.
- CZAYKOWSKA-HIGGINS, EWA. 2009. Research models, community engagement, and linguistic fieldwork: reflections on working within Canadian Indigenous communities. *Language Documentation and Conservation* 3(1). 15–50. Online: <http://hdl.handle.net/10125/4423>.
- DAWSON, CHARLES A. 1936. *Group settlement: ethnic communities in Western Canada*. Toronto, Ontario: Macmillan.
- DI PAOLO, MARIANNA and MALCAH YAEGER-DROR (eds.). 2010. *Sociophonetics: a student's guide*. London: Routledge.
- DOELL, CYNTHIA R. n.d. *A Plautdietsch grammar, proverbs, stories and glossary*. Unpublished thesis, University of Nebraska.
- DOELL, LEONARD. 1987. *The Bergthaler Mennonite Church of Saskatchewan, 1892–1975*. Winnipeg, Manitoba: CMBC Publications.
- DOELL, LEONARD. 2001. Swift Current Old Colony Mennonite Church. In PLETT, DELBERT F. (ed.), *Old Colony Mennonites in Canada: 1875 to 2000*, 152–155. Steinbach, Manitoba: Crossway Publications.
- DORIAN, NANCY C. 2010. *Investigating variation: the effects of social organization and social setting*. Oxford: Oxford University Press.
- DRIEDGER, JACOB M. 2011. *Enn Bloomenheim Oppjewossen* [Growing up in Blumenheim]. Altona,



Manitoba: Friesens.

- DUBENION-SMITH, SHANNON A. 2010. Verb complex phenomena in West Central German: empirical domain and multi-causal account. *Journal of Germanic Linguistics* 22(2). 99–191.
- DU BOIS, JOHN W., WALLACE L. CHAFE, CHARLES MEYER, SANDRA A. THOMPSON, ROBERT ENGLEBRETSON, and NII MARTEY. 2000–2005. *Santa Barbara corpus of spoken American English, Parts 1–4*. Philadelphia: Linguistic Data Consortium.
- DWYER, ARIENNE M. 2006. Ethics and practicalities of cooperative fieldwork and analysis. In GIPPERT, JOST, NIKOLAUS P HIMMELMANN, and ULRIKE MOSEL (eds.), *Essentials of Language Documentation*, 31–66. Berlin: Mouton de Gruyter.
- DYCK, HENRY DIETRICH. 1964. *Language differentiation in two Low German groups in Canada*. Unpublished doctoral dissertation, University of Pennsylvania.
- DYCK, CORNELIUS J. 1993. *An introduction to Mennonite history. 3rd edn*. Scottdale, Pennsylvania: Herald Press.
- EDMONT, EDMOND and JULES GILLIÉRON. 1902. *Atlas linguistique de la France* [Linguistic atlas of France]. Paris: Honoré Champion.
- ENS, ADOLF. 1994. *Subjects or citizens? The Mennonite experience in Canada, 1870–1925*. Ottawa, Ontario: University of Ottawa Press.
- EPP, FRANK H. 1962. *Mennonite exodus: the rescue and resettlement of the Russian Mennonites since the Communist Revolution*. Altona, Manitoba: D.W. Friesen & Sons.
- EPP, FRANK H. 1974. *Mennonites in Canada, 1786–1920: the history of a separate people*. Toronto, Ontario: Macmillan.
- EPP, FRANK H. 1982. *Mennonites in Canada, 1920–1940: a people's struggle for survival*. Toronto, Ontario: Macmillan.
- EPP, J. H., MARLENE EPP, and RICHARD D. THIESSEN. 2009. Brotherfield Mennonite Brethren Church (Waldheim, Saskatchewan, Canada). *Global Anabaptist Mennonite Encyclopedia Online*. Online: <http://www.gameo.org/encyclopedia/contents/B767ME.html>
- EPP, REUBEN. 1972. *Plautdietsche Schreftsteckja (Low-German writings)*. Steinbach, Manitoba: Derksen Printers.
- EPP, REUBEN. 1987. Plautdietsch: origins, development and state of the Mennonite Low German language. *Journal of Mennonite Studies* 5.61–72.

- EPP, REUBEN. 1993. *The story of Low German & Plautdietsch: tracing a language across the globe*. Hillsboro, Kansas: Reader's Press.
- EPP, REUBEN. 1996. *The spelling of Low German & Plautdietsch: toward an official Plautdietsch orthography*. Hillsboro, Kansas: Reader's Press.
- EPP, REUBEN. 2000. *161 Plautdietsch verbs fully conjugated in all tenses and a writing system*. Kelowna, British Columbia: unpublished ms.
- EVANS, NICHOLAS. 2010. *Dying words: endangered languages and what they have to tell us*. Malden, Massachusetts: Wiley-Blackwell.
- EVANS, NICHOLAS and ALAN DENCH. 2006. Introduction: catching language. In AMEKA, FELIX K., ALAN DENCH, and NICHOLAS EVANS (eds.), *Catching language: the standing challenge of grammar writing*, 1–39. Berlin / New York: Mouton de Gruyter.
- EVERS, ARNOLD. 1975. *The transformational cycle of Dutch and German*. Doctoral dissertation, University of Utrecht.
- FEHR, JACOB M. 1993. *De Stuk Maschien* [The manual washing machine]. Swift Current, Saskatchewan: n.p.
- FEHR, JACOB M. 2001. *Knak Sot* [Sunflower seeds]. Steinbach, Manitoba: Derksen Printers.
- FEHR, JACOB M. 2006. *Plautdietsche Bibel Jeschichten fa Kjinja* [Plautdietsch Bible stories for children]. Steinbach, Manitoba: Derksen Printers.
- FLORES FARFÁN, JOSÉ ANTONIO and FERNANDO F. RAMALLO (eds.). 2010. *New perspectives on endangered languages: bridging gaps between sociolinguistics, documentation and language revitalization*. Amsterdam: John Benjamins.
- FRIESEN, RICHARD J. 1975. *Old Colony Mennonite settlements in Saskatchewan: a study in settlement change*. Unpublished MA thesis, University of Alberta.
- FRIESEN, VICTOR CARL. 1988. *The windmill turning: nursery rhymes, maxims, and other expressions of Western Canadian Mennonites*. Edmonton, Alberta: University of Alberta Press.
- GLENDINNING, LESLEY. 2006. *Mennonites at play: postmodern aspects of Low German drama*. Unpublished MA thesis, University of Manitoba.
- GLYNN, DYLAN. 2009. Polysemy, syntax, and variation: a usage-based method for cognitive semantics. *New Directions in Cognitive Linguistics* 24. 77–104.

- GLYNN, DYLAN. 2010. Testing the hypothesis: objectivity and verification in usage-based cognitive semantics. In GLYNN, DYLAN and KERSTIN FISCHER (eds.), *Corpus-driven cognitive semantics. Quantitative approaches*, 239–270. Berlin: Mouton de Gruyter.
- GOEBL, HANS. 1982. *Dialektometrie: Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie* [Dialectometry: principles and methods for the application of numerical taxonomy in the area of dialect geography]. Wien: Österreichische Akademie der Wissenschaften.
- GOERZEN, JAKOB WARKENTIN. 1950. *The phonology of Plautdietsch*. Unpublished MA thesis, University of Alberta.
- GOERZEN, JAKOB WARKENTIN. 1972. *Low German in Canada: a study of "Plautdietsch."* Edmonton, Alberta: n.p.
- GUENTER, JACOB G. 1981. *Men of Steele: life style of a unique sect*. Saskatoon, Saskatchewan: n.p.
- GUENTER, JACOB G., LEONARD DOELL, DICK BRAUN, JACOB L. GUENTHER, HENRY A. FRIESEN, JACOB W. LOEPPKY, JOHN P. DOELL, PETER G. GIESBRECHT, ANNA BUECKERT, ANNA BRAUN, and MARY GOERTZEN (eds.). 1995. *Hague-Osler Mennonite Reserve: 1895–1995*. Saskatoon, Saskatchewan: Hague-Osler Reserve Book Committee.
- HAARMANN, HARALD. 2005. Roofless dialects / Dachlose Dialekte. In DITTMAR, NORBERT, KLAUS J. MATTHEIER, and ULRICH AMMON (eds.), *Sociolinguistics (Soziolinguistik): an international handbook of the science of language and society (ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft)*. 2nd edn., 1545–1551. Berlin: Mouton de Gruyter.
- HARRISON, K. DAVID. 2007. *When languages die: the extinction of the world's languages and the erosion of human knowledge*. Oxford: Oxford University Press.
- HARTMAN KEISER, STEVEN. 2009. When 'speech islands' aren't islands: parallel independent development, drift, and minimal levels of contact for diffusion. *Diachronica* 26(1). 1–35.
- HEDGES, KELLY LYNN. 1996. *Plautdietsch and Huuchdietsch in Chihuahua: language, literacy, and identity among the Old Colony Mennonites in northern Mexico*. Unpublished doctoral dissertation, Yale University.
- HEPP, MICHAEL J. 1998. *The legacy of St. Peter's Colony*. Muenster, Saskatchewan: St. Peter's Press. Online: <http://www.ourroots.ca/e/toc.aspx?id=6058>

- HILL, JANE H. 2002. "Expert rhetorics" in advocacy for endangered languages: who is listening, and what do they hear? *Journal of Linguistic Anthropology* 12(2). 119–133.
- HIMMELMANN, NIKOLAUS P. 1998. Documentary and descriptive linguistics. *Linguistics* 36. 161–195.
- HIMMELMANN, NIKOLAUS P. 2006. Language documentation: what is it and what is it good for? In GIPPERT, JOST, NIKOLAUS P HIMMELMANN, and ULRIKE MOSEL (eds.), *Essentials of language documentation*, 1–30. Berlin: Mouton de Gruyter.
- HIMMELMANN, NIKOLAUS P. 2008. Reproduction and preservation of linguistic knowledge: linguistics' response to language endangerment. *Annual Review of Anthropology* 37. 337–350.
- HIMMELMANN, NIKOLAUS P. 2012. Linguistic data types and the interface between language documentation and description. *Language Documentation and Conservation* 6. 187–207. Online: <http://hdl.handle.net/10125/4503>.
- HUSSON, FRANÇOIS, JULIE JOSSE, and JÉRÔME PAGÈS. 2010. Principal component methods – hierarchical clustering – partitional clustering: why would we need to choose for visualizing data? *Technical Report - Agrocampus, Applied Mathematics Department*, 1–17.
- JANZEN, ABRAM G. 2004. *Memories of an Old Colony Waisenamt manager*. Hague, Saskatchewan: n.p.
- JEDIG, HUGO H. 1966. *Laut- und Formenbestand der niederdeutschen Mundart des Altai-Gebietes* [Phonological and morphological inventory of the Low German dialect of the Altai region]. *Sitzungsberichte der sächsischen Akademie der Wissenschaften zu Leipzig, Phil.-histor. Klasse* 112 (5). Berlin: Akademie-Verlag.
- JEDIG, HUGO H. 1969. *Очерки по синтаксису нижненемецкого говора алтайского края* [Studies of the syntax of the Low German language of the Altai region]. Omsk: западно-сибирское книжное издательство.
- JOHANNESSEN, JANNE BONDI. 2011. The Nordic Dialect Corpus: a joint research infrastructure. *Oslo Studies in Language* 3(2). 45–62.
- KANAKIN, IGOR and MARGARITA WALL. 1994. *Das Plautdietsch in Westsibirien* [Plautdietsch in western Siberia]. *Lingua Mennonitica*. Groningen: Rijksuniversiteit Groningen.

- KAUFMANN, GÖZ. 1997. *Varietätendynamik in Sprachkontaktsituationen: Attitüden und Sprachverhalten rußlanddeutscher Mennoniten in Mexiko und den USA* [Variety dynamics in language contact situations: attitudes and linguistic behaviour of Russian-German Mennonites in Mexico and the USA]. Frankfurt am Main: Peter Lang Verlag.
- KAUFMANN, GÖZ. 2003a. Des Plattdeutschen Wanderjahre oder die lexikalischen Folgen der mennonitischen Flucht nach Amerika [Low German's years of wandering, or the lexical consequences of the Mennonite flight to America]. In ANDROUTSOPOULOS, JANNIS and EVELYN ZIEGLER (eds.), *“Standardfragen”: Soziolinguistische Perspektiven auf Sprachgeschichte, Sprachkontakt, und Sprachvariation. Variolinguia 18*, 139–160. Frankfurt am Main: Peter Lang Verlag.
- KAUFMANN, GÖZ. 2003b. The verb cluster in Mennonite Low German. In MATTHEIER, KLAUS J. and WILLIAM D. KEEL (eds.), *German language varieties worldwide: internal and external perspectives / Deutsche Sprachinseln weltweit: interne und externe Perspektive*, 177–198. Frankfurt am Main: Peter Lang Verlag.
- KAUFMANN, GÖZ. 2005. Der eigensinnige Informant: Ärgernis bei der Datenerhebung oder Chance zum analytischen Mehrwert. [The wilful informant: a nuisance for data collection or an opportunity for added value in analysis.] In LENZ, FRIEDRICH and STEFAN J. SCHIERHOLZ (eds.), *Corpuslinguistik in Lexik und Grammatik*, 61–95. Tübingen: Stauffenberg Verlag.
- KAUFMANN, GÖZ. 2007. The verb cluster in Mennonite Low German: a new approach to an old topic. *Linguistische Berichte* 210. 147–207.
- KAUFMANN, GÖZ. 2008. Where syntax meets morphology: Varianten des bestimmten Artikels und die Variation satzfinaler Verbcluster im Plattdeutschen texanischer Mennoniten [Variants of the definite article and the variation of sentence-final verb clusters in the Low German of Texas Mennonites]. In PATOCKA, FRANZ and GUIDO SEILER (eds.), *Dialektale Morphologie, dialektale Syntax. Beiträge zum 2. Kongress der Internationalen Gesellschaft für Dialektologie des Deutschen*. Vienna: Praesens Verlag.
- KEEL, WILLIAM D. 2006. Plattdүүtsch and Plautdietsch in western Missouri and Kansas: the resilience of Low German networks on the Great Plains. In FULLER, JANET M. and LINDA L. THORNBURG (eds.), *Studies in contact linguistics: essays in honor of Glenn G. Gilbert*, 147–166. New York: Peter Lang Verlag.

- KENDALL, TYLER. 2011. Corpora from a sociolinguistic perspective. In GRIES, STEFAN TH. (ed.), *Corpus studies: future directions. Special issue of Revista Brasileira de Linguística Aplicada* 11(2). 361–389.
- KLASSEN, HEINRICH N. 1969. Untersuchungen zum grammatischen Bau der niederdeutschen Mundart im Gebiet Orenburg (RSFSR) [Investigations of the grammatical structure of the Low German dialect in the region of Orenburg (USSR)]. *Wissenschaftliche Zeitschrift der Martin-Luther-Universität Halle-Wittenberg, Gesellschafts- und sprachwissenschaftliche Reihe* 18(6). 27–48.
- KLOSS, HEINZ. 1989. Sociolinguistic parallels between the Mennonite speakers of Pennsylvania German (or Pennsylvanisch) and of Plautdietsch. In ENNINGER, WERNER, JOACHIM RAITH, and KARL-HEINZ WANDT (eds.), *Studies on the languages and the verbal behavior of the Pennsylvania Germans II. Zeitschrift für Dialektologie und Linguistik, Beiheft 64*, 117–124. Stuttgart: Franz Steiner Verlag.
- KRAHN, CORNELIUS. 1959. Plattdeutsch [Low German]. In KRAHN, CORNELIUS et al. (eds.), *The Mennonite Encyclopedia*. Scottsdale, Pennsylvania: Mennonite Publishing House.
- KRETZSCHMAR, WILLIAM A., JR., JEAN ANDERSON, JOAN C. BEAL, KAREN P. CORRIGAN, LISA LENA OPAS-HÄNNINEN and BARTŁOMIEJ PŁICHTA. 2006. Collaboration on corpora for regional and social analysis. *Journal of English Linguistics* 34(3). 172–205.
- LABOV, WILLIAM. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania.
- LABOV WILLIAM, SHARON ASH, and CHARLES BOBERG. 2006. *The atlas of North American English: phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.
- LADEFOGED, PETER. 2003. *Phonetic data analysis: an introduction to fieldwork and instrumental techniques*. Malden, Massachusetts: Blackwell.
- LAMELI, ALFRED. 2010. Linguistic atlases – traditional and modern. In AUER, PETER and JÜRGEN ERICH SCHMIDT (eds.), *Language and space: an international handbook of linguistic variation. Volume 1: theories and methods*, 567–591. Berlin: Mouton de Gruyter.
- LASCH, AGATHE. 1914. *Mittelniederdeutsche Grammatik* [Middle Low German grammar]. Halle: Max Niemeyer.
- LEE, NALA HUIYING. 2013. Review of Shure WH30XLR cardioid headset microphone and Countryman E6 omnidirectional earset microphone. *Language Documentation and*

- Conservation* 7: 177–184. Online: <http://hdl.handle.net/10125/4595>
- LEECH, GEOFFREY. 1992. Corpora and theories of linguistic performance. In: SVARTVIK, JAN (ed.), *Directions in corpus linguistics: proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*, 105–125. Berlin: Mouton de Gruyter.
- LEHN, WALTER ISAAK. 1957. *Rosenthal Low German: Synchronic and diachronic phonology*. Unpublished doctoral dissertation, Cornell University.
- LEONARD, WESLEY Y. and ERIN HAYNES. 2010. Making “collaboration” collaborative: an examination of perspectives that frame linguistic field research. *Language Documentation and Conservation* 4. 269–293.
- LOEWEN, HARRY and AL REIMER. 1985. Origins and literary development of Canadian-Mennonite Low German. *Mennonite Quarterly Review* 59(3). 279–286.
- LOEWEN, JACOB A. 1996. *Onze iashte Missjounszreiz* [Our first missions trip]. Winnipeg, Manitoba: Mennonite Books.
- LOEWEN, JACOB A. 1998. Toward a phonemic alphabet of Plautdietsch. *Journal of Mennonite Studies* 16. 128–146.
- LÜPKE, FRIEDERIKE. 2009. Data collection methods for field-based language documentation. In AUSTIN, PETER K. (ed.), *Language Documentation and Description. Volume 6*, 53–100. London: School of Oriental and African Studies.
- MALKIEL, YAKOV. 1976. From Romance philology through dialect geography to sociolinguistics. *International Journal of the Sociology of Language* 9, 59–84.
- MATTHEIER, KLAUS J. 1994. Theorie der Sprachinsel: Voraussetzungen und Strukturierungen [Linguistic enclave theory: suppositions and structurings]. In BEREND, NINA and KLAUS J. MATTHEIER (eds.), *Sprachinselforschung: Eine Gedenkschrift für Hugo Jedig*, 333–348. Frankfurt am Main: Peter Lang Verlag.
- MCCAFFERY, ISAIAS J. 2008. *Wi leahre Plautdietsch: a beginner’s guide to Mennonite Low German*. Morrisville, North Carolina: Lulu.com.
- MCENERY, TONY and NICK OSTLER. 2000. A new agenda for corpus linguistics: working with all of the world’s languages. *Literary and Linguistic Computing* 15 (4). 403–419.
- MCENERY, TONY, RICHARD XIAO, and YUKIO TONO. 2006. *Corpus-based language studies: an advanced resource book*. New York: Routledge.

- McIVER, VERA. 1996. *Quest for freedom*. Victoria, British Columbia: Trafford.
- McLENNAN, DAVID. 2008. *Our towns: Saskatchewan communities from Abbey to Zenon Park*. Regina, Saskatchewan: University of Regina Press.
- MEIER, JÜRGEN. 1978. Zur Behandlung der Syntax in den niederdeutschen Dialektwörterbüchern [On the treatment of syntax in Low German dialect dictionaries.] *Zeitschrift für Dialektologie und Linguistik* 45 (3). 289–311.
- MEYER, CHARLES F. 2004. Can you really study language variation in linguistic corpora? *American Speech* 79(4). 339–355.
- MEYERHOFF, MIRIAM and NAOMI NAGY. 2008. Introduction: social lives in language. In MEYERHOFF, MIRIAM and NAOMI NAGY (eds.), *Social lives in language: the sociolinguistics of multilingual speech communities. Celebrating the work of Gillian Sankoff*, 1–16. Amsterdam / Philadelphia: John Benjamins.
- MIERAU, ERIC. 1964. *A descriptive grammar of Ukrainian Low German*. Unpublished doctoral dissertation, Indiana University.
- MILROY, LESLEY. 1987. *Observing and analyzing natural language: a critical account of sociolinguistic method*. Oxford: Blackwell.
- MITZKA, WALTHER. 1930. Die Sprache der deutschen Mennoniten [The language of the German Mennonites]. *Heimatblätter des Deutschen Heimatbundes* 8(1). Danzig: A. W. Kafemann.
- MITZKA, WALTHER. 1968a [1922]. Dialektgeographie der Danziger Nehrung [Dialect geography of the Danzig Spit]. In SCHMITT, LUDWIG ERICH (ed.), *Kleine Schriften zur Sprachgeschichte und Sprachgeographie*, 191–210. Berlin: Walter de Gruyter.
- MITZKA, WALTHER. 1968b [1924]. Sprachgeschichtliche Streifzüge auf der Danziger Nehrung [Historical linguistic forays on the Danzig Spit]. In SCHMITT, LUDWIG ERICH (ed.), *Kleine Schriften zur Sprachgeschichte und Sprachgeographie*, 211–230. Berlin: Walter de Gruyter.
- MOELLEKEN, WOLFGANG W. 1966. Low German in Mexico. *Publications of the American Dialect Society* 46. 31–39.
- MOELLEKEN, WOLFGANG W. 1967. Diaphonic correspondences in the Low German of Mennonites from the Fraser Valley, British Columbia. *Zeitschrift für Mundartforschung* 34(3). 240–



- MOELLEKEN, WOLFGANG W. 1972. *Niederdeutsch der Molotschna- und Chortitzamennoniten in British Columbia / Kanada* [Low German of the Molochnaya and Chortitza Mennonites in British Columbia / Canada]. *Phonai 10. Lautbibliothek der europäischen Sprachen und Mundarten. Deutsche Reihe. Monographien 4*. Tübingen: Max Niemeyer Verlag.
- MOELLEKEN, WOLFGANG W. 1987. Die linguistische Heimat der russlanddeutschen Mennoniten in Kanada und Mexiko [The linguistic home of the Russian German Mennonites in Canada and Mexico]. *Niederdeutsches Jahrbuch* 110. 89–123.
- MOELLEKEN, WOLFGANG W. 1993. Retroflexes r im Niederdeutsch der mexikanischen Altkolonier Mennoniten [Retroflex r in the Low German of the Mexican Old Colony Mennonites]. *Zeitschrift für Dialektologie und Linguistik* 60(1). 21–38.
- MOELLEKEN, WOLFGANG W. 1994. The demise of Mennonite Low German in Reedley, California. In BEREND, NINA and KLAUS J. MATTHEIER (eds.), *Sprachinselforschung: eine Gedenkschrift für Hugo Jedig*, 303–317. Frankfurt am Main: Peter Lang Verlag.
- MOISL, HERMANN. 2009. Using electronic corpora to study language variation: the problem of data sparsity. In TSIPLAKOU, STAVROULA, MARILENA KARYOLEMOU, and PAVLOS PAVLOU (eds.), *Language variation—European perspectives II. Selected papers from the 4th International Conference on Language Variation in Europe (ICLAVE 4), Nicosia, June 2007*, 169–178. Amsterdam / Philadelphia: John Benjamins.
- MULDER, JEAN and HOLLY SELLERS. 2010. Classifying clitics in Sm’algyax: approaching theory from the field. In BEREZ, ANDREA L., JEAN MULDER, and DAISY ROSENBLUM (eds.), *Fieldwork and linguistic analysis in Indigenous languages of the Americas*. Language Documentation & Conservation Special Publication No. 2, 33–56. Honolulu: University of Hawai’i Press.
- NAGY, NAOMI. 2012. Sociolinguistics and phonology. In BAYLEY, ROBERT, RICHARD CAMERON, and CEIL LUCAS (eds.), *Oxford handbook of sociolinguistics*, 425–444. Oxford: Oxford University Press.
- NAGY, NAOMI and MIRIAM MEYERHOFF. 2013. Extending ELAN into quantitative sociolinguistics. Poster presented at the third International Conference on Language Documentation and Conservation, February 28–March 3, University of Hawai’i at Mānoa.

- NAIDITSCH, LARISSA. 2001. Die Entwicklung des Vokalismus im Mennonitenplatt [The development of vowels in Mennonite Low German]. In BRAUN, ANGELIKA (ed.), *Beiträge zu Linguistik und Phonetik. Festschrift für Joachim Göschel zum 70. Geburtstag*, 242–256. Stuttgart: Franz Steiner Verlag.
- NATHAN, DAVID. 2006. Thick interfaces: mobilizing language documentation with multimedia. In GIPPERT, JOST, NIKOLAUS P HIMMELMANN, and ULRIKE MOSEL (eds.), *Essentials of Language Documentation*, 363–379. Berlin: Mouton de Gruyter.
- NELSON, GERALD. 2002. *International Corpus of English: markup manual for spoken texts*. Unpublished ms. Online: <http://ice-corpora.net/ice/spoken.doc>
- NEUFELD, ELDO. 2000a. *Plautdietsch grammar. an aid to speaking, reading, and writing Netherlandic-Mennonite Plautdietsch. LINCOS Studies in Germanic Linguistics 8*. Munich: LINCOS Europa.
- NEUFELD, ELDO. 2000b. *Plautdietsch verb conjugation. Volume I. LINCOS Studies in Germanic Linguistics 9*. Munich: LINCOS Europa.
- NEUFELD, ELDO. 2000c. *Plautdietsch verb conjugation. Volume II. LINCOS Studies in Germanic Linguistics 10*. Munich: LINCOS Europa.
- NERBONNE, JOHN and WILBERT HEERINGA. 2001. Computational comparison and classification of dialects. *Dialectologia et Geolinguistica* 9. 69–83.
- NERBONNE, JOHN, RINKE COLEN, CHARLOTTE GOOSKENS, PETER KLEIWEG, and THERESE LEINONEN. 2011. Gabmap – a web application for dialectology. *Dialectologia: Revista electrònica*. Special Issue II. 65–89.
- NEWMAN, JOHN. 2008. Spoken corpora: rationale and application. *Taiwan Journal of Linguistics* 6(2). 27–58.
- NIEBAUM, HERMANN. 2000. Phonetik und Phonologie, Graphetik und Graphemik des Mittelniederdeutschen [Phonetics and phonology, graphetics and graphemics of Middle Low German]. In BESCH, WERNER, ANNE BETTEN, OSKAR REICHMANN, and STEFAN SONDEREGGER (eds.), *Sprachgeschichte: Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. Zweiter Teilband* (2nd rev. edn.), 1422–1430. Berlin: Mouton de Gruyter.
- NIEUWEBOER, ROGIER. 1998. *The Altai dialect of Plautdiitsch (West-Siberian Mennonite Low*

- German*). Groningen: Rijksuniversiteit Groningen.
- NIEUWEBOER, ROGIER and TJEERD DE GRAAF. 1994. The language of the West Siberian Mennonites. *Internationalt tidsskrift for sprog og kommunikation* 1. 47–61.
- NYMAN, LENNART. 1978. Die niederdeutschen Mundarten der Rußland-Mennoniten [The Low German dialects of the Russian Mennonites]. In HYLDEGAARD-JENSEN, KARL (ed.), *Beiträge zur historischen Grammatik und zur Dialektologie. Kopenhagener Beiträge zur Germanistischen Linguistik* 10, 43–56. Copenhagen: Akademisk.
- NYMAN, LENNART. 1997. Einige Beobachtungen zu Varietäten und Varietätenausgleich im Niederdeutschen der Orenburger Rußland-Mennoniten [Some observations on varieties and variety levelling in the Low German of the Orenburg Russian Mennonites]. In BRANDT, GISELA (ed.), *Historische Soziolinguistik des Deutschen III. Sprachgebrauch und sprachliche Leistung in sozialen Schichten und soziefunktionalen Gruppen. Internationale Fachtagung, Rostock / Kühlingsborn, 15.-18.9.1996. Stuttgarter Arbeiten zur Germanistik* 351, 261–276. Stuttgart: Hans-Dieter Heinz Verlag.
- PAPROSKI, PAUL. 2011. *The German Catholics of St. Peter's Colony: 1903–1930*. Unpublished MA thesis, University of Saskatchewan.
- PLETT, DELBERT F. 2000. 'Poor and simple?' The economic background of the Mennonite immigrants to Manitoba, 1874–79. *Journal of Mennonite Studies* 18. 114–128.
- PLETT, DELBERT F. 2001. 'Poor and simple?' The economic background of the 1874 Russian Mennonite emigration to Manitoba. In PLETT, DELBERT F. (ed.), *Old Colony Mennonites in Canada: 1875 to 2000*, 25–26. Steinbach, Manitoba: Crossway Publications.
- POPLACK, SHANA. 1989. The care and handling of a mega-corpus: the Ottawa-Hull French Project. In FASOLD, RALPH W. and DEBORAH SCHIFFRIN (eds.), *Language Change and Variation*, 411–451. Amsterdam: John Benjamins.
- PRESTON, DENNIS R. 1985. The Li'l Abner syndrome: written representations of speech. *American Speech* 60 (4), 328–336.
- PRESTON, DENNIS R. 1989. *Perceptual dialectology: nonlinguists' views of areal linguistics*. Dordrecht: Foris.
- PRESTON, DENNIS R. (ed.). 1999. *Handbook of perceptual dialectology. Volume 1*. Amsterdam: John Benjamins.

- PRIETO, PILAR, TERESA CABRÉ, and MARIA DEL MAR VANRELL. 2010. L'Atles interactiu de l'entonació del català i el traçat de les isoglosses entonatives del català [The interactive atlas of Catalan intonation and the treatment of the intonational isoglosses of Catalan]. *Caplletra* 49. 201–226.
- PROKIĆ, JELENA, ÇAĞRI ÇÖLTEKIN, and JOHN NERBONNE (2012). Detecting shibboleths. In BUTT, MIRIAM and JELENA PROKIĆ (eds.), *Visualization of language patterns and uncovering language history from multilingual resources*, 72–80. Workshop at the 13th Conference of the European Chapter of the Association for Computational Linguistics. Avignon, France.
- PUTNAM, MICHAEL T. (ed.) 2011. *Studies on German-language islands*. Amsterdam: John Benjamins.
- QUIRING, JACOB. 1928. *Die Mundart von Chortitza in Süd-Rußland* [The dialect of Chortitza in Southern Russia]. Munich: Druckerei Studentenheim München.
- REDEKOP, CALVIN. 1969. *The Old Colony Mennonites: dilemmas of ethnic minority life*. Baltimore, Maryland: Johns Hopkins Press.
- REGHEHR, T. D. 1996. *Mennonites in Canada, 1939–1970: a people transformed*. Toronto, Ontario: University of Toronto Press.
- REGER, ADINA and DELBERT PLETT. 2001. *Diese Steine: Die Russlandmennoniten* [These stones: the Russian Mennonites]. Steinbach, Manitoba: Crossway Publications.
- REIMER, AL. 1982. There's now an 'official' way to write Low German. *Mennonite Mirror* 11 (June). 7–8.
- REIMER, AL. 1984. "Derche Bloom Råde": Arnold Dyck and the comic irony of the *Forstei*. *Journal of Mennonite Studies* 2(1). 60–71.
- REIMER, AL. 1991. The role of Arnold Dyck in Canadian Mennonite writing. *Journal of Mennonite Studies* 9(1). 83–91.
- REIMER, AL, ANNE REIMER, and JACK THIESSEN (eds.). 1983. *A Sackful of Plautdietsch*. Winnipeg, Manitoba: Hyperion Press.
- REMPEL, HERMAN. 1995. *Kjenn jie noch Plautdietsch? A Mennonite Low German dictionary*. 2nd rev. edn. Rosenort, Manitoba: PrairieView Press.
- RIEHL, CLAUDIA MARIA. 2010. Discontinuous language spaces (Sprachinseln). In AUER, PETER and

- JÜRGEN ERICH SCHMIDT (eds.), *Language and space: an international handbook of linguistic variation. Volume 1: theories and methods*, 332–354. Berlin: Mouton de Gruyter.
- ROHKOHL, KAI. 1993. *Die plautdietsche Sprachinseln Fernheim/Chaco (Paraguay): Dokumentation des Sprachverhaltens einer russlanddeutschen Mennonitenkolonie* [The Low German linguistic enclave Fernheim/Chaco (Paraguay): documentation of the linguistic behaviour of a Russian-German Mennonite colony]. Schriftenreihe der Kommission für ostdeutsche Volkskunde in der Deutschen Gesellschaft für Volkskunde e.V. 64. Marburg: N. G. Elwert Verlag.
- SALTVEIT, LAURITS. 1983. Syntax. In CORDES, GERHARD and DIETER MÖHN (eds.), *Handbuch zur niederdeutschen Sprach- und Literaturwissenschaft*, 279–333. Berlin: Erich Schmidt Verlag.
- SANKOFF, DAVID. 2005. Problems of representativeness / Probleme der Repräsentativität. In DITTMAR, NORBERT, KLAUS J. MATTHEIER, and ULRICH AMMON (eds.), *Sociolinguistics (Soziolinguistik): an international handbook of the science of language and society (ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft)*. 2nd edn., 998–1002. Berlin: Mouton de Gruyter.
- SAPP, CHRISTOPHER D. 2006. *Verb order in subordinate clauses from Early New High German to Modern German*. Unpublished doctoral dissertation, Indiana University.
- SCHAPANSKY, HENRY. 2006. *Mennonite migrations (and the Old Colony)*. Rosenort, Manitoba: self published.
- SCHIRMUNSKI, VIKTOR. 1928. *Die deutschen Kolonien in der Ukraine* [The German colonies in the Ukraine]. Moscow: Zentral-Völkerverlag der Sowjet-Union.
- SCHRAUF, ROBERT W. 2013. Using correspondence analysis to model immigrant multilingualism over time. In DUARTE, JOANA and INGRID GOGOLIN (eds.), *Linguistic superdiversity in urban areas: research approaches*, 27–44. Amsterdam / Philadelphia: John Benjamins.
- SCHROEDER, WILLIAM and HELMUT T. HUEBERT. 1996. *Mennonite historical atlas*. 2nd edn. Winnipeg, Manitoba: Springfield Publishers.
- SCHULTZE-BERNDT, EVA. 2006. Linguistic annotation. In GIPPERT, JOST, NIKOLAUS P HIMMELMANN, and ULRIKE MOSEL (eds.), *Essentials of language documentation*, 213–251. Berlin: Mouton de

Gruyter.

- SEIFART, FRANK, GEOFFREY HAIG, NIKOLAUS P. HIMMELMANN, DAGMAR JUNG, ANNA MARGETTS, and PAUL TRILSBEEK (eds.). 2012. *Potentials of language documentation: methods, analyses, and utilization*. Language Documentation and Conservation Special Publication No. 3. Honolulu: University of Hawai'i Press.
- SÉGUY, JEAN. 1971. La relation entre la distance spatiale et la distance lexicale [The relationship between spatial distance and lexical distance]. *Revue de Linguistique Romane* 35. 335–357.
- SIEMENS, HEINRICH. 2012. *Plautdietsch: Grammatik, Geschichte, Perspektiven* [Plautdietsch: grammar, history, perspectives]. Bonn: Tweeback Verlag.
- SPRUIT, MARCO RENÉ. 2008. *Quantitative perspectives on syntactic variation in Dutch dialects*. Utrecht: Landelijke Onderzoekschool Taalwetenschap.
- STANFORD, JAMES N. 2012. One size fits all? Dialectometry in a small clan-based indigenous society. *Language Variation and Change* 24. 247–278.
- STANFORD, JAMES N. and DENNIS R. PRESTON. 2009. The lure of a distant horizon: variation in Indigenous minority languages. In STANFORD, JAMES N. and DENNIS R. PRESTON (eds.), *Variation in Indigenous minority languages*, 1–20. Amsterdam: John Benjamins.
- STEFFEN, JOACHIM. 2006a. *Vereinzelte Sprachinseln oder Archipel? Die Mennonitenkolonien in Belize im englisch-spanischen Sprachkontakt* [Isolated language islands or archipelago? The Mennonite colonies in Belize in English-Spanish language contact]. Band 1: Textband. Kiel: Westensee-Verlag.
- STEFFEN, JOACHIM. 2006b. *Vereinzelte Sprachinseln oder Archipel? Die Mennonitenkolonien in Belize im englisch-spanischen Sprachkontakt* [Isolated language islands or archipelago? The Mennonite colonies in Belize in English-Spanish language contact]. Band 2: Karteband. Kiel: Westensee-Verlag.
- STROBL, CAROLIN, JAMES MALLEY, and GERHARD TUTZ. 2009. An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests. *Ludwig-Maximilians-Universität München Department of Statistics Technical Reports* 55. 1–42. Online: <http://epub.ub.uni-muenchen.de/10589/1/partitioning.pdf>

- STUBBS, MICHAEL. 2001. *Words and phrases: corpus studies of lexical semantics*. Oxford: Blackwell.
- SUDERMAN, ELMER F. 1969. The comic spirit of Arnold Dyck. *Mennonite Life* (October 1969). 169–170.
- SZMRECSANYI, BENEDIKT. 2011. Corpus-based dialectometry: a methodological sketch. *Corpora* 6(1). 45–76.
- SZMRECSANYI, BENEDIKT. 2014. Forests, trees, corpora, and dialect grammars. In SZMRECSANYI, BENEDIKT and BERNHARD WÄLCHLI (eds.), *Aggregating dialectology, typology, and register analysis: linguistic variation in text and speech*, 89–112. Berlin: de Gruyter.
- TAGLIAMONTE, SALI A. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.
- TAGLIAMONTE, SALI A. and R. HARALD BAAYEN. 2012. Models, forests, and trees of York English: *was/were* variation as a case study for statistical practice. *Language Variation and Change* 24. 135–178.
- TEUBERT, WOLFGANG. 2001. Corpus linguistics and lexicography: more than a slogan? *International Journal of Corpus Linguistics* 6. 125–153.
- THIEBERGER, NICHOLAS and ANDREA L. BEREZ. 2012. Linguistic data management. In THIEBERGER, NICHOLAS (ed.), *The Oxford handbook of linguistic fieldwork*, 90–118. Oxford: Oxford University Press.
- THIESSEN, JOHN [JACK]. 1963. *Studien zum Wortschatz der kanadischen Mennoniten* [Studies of the lexicon of the Canadian Mennonites]. Deutsche Dialektgeographie Band 64. Marburg: N. G. Elwert Verlag.
- THIESSEN, JACK. 1989. A new look at an old problem: origins of the variations in Mennonite Plautdietsch. *Mennonite Quarterly Review* 63(3). 285–296.
- THIESSEN, JACK. 2006. Dutch-Flemish words in Mennonite Low German. *Preservings: A Journal of the D. F. Plett Historical Research Foundation* 26. 23–26.
- TOLKSDORF, ULRICH. 1985. Die Mundarten Danzigs und seines Umlandes [The dialects of Danzig and its surroundings]. In JÄHNIG, BERNHART and PETER LETKEMANN (eds.), *Danzig in acht Jahrhunderten: Beiträge zur Geschichte eines hansischen und preussischen Mittelpunktes*, 313–336. Münster: Nicolaus-Copernicus-Verlag.

- TRUDGILL, PETER. 1999. The chaos before the order: New Zealand English and the second stage of new dialect formation. In Jahr, Ernst Håkon (ed.), *Language change: advances in historical sociolinguistics*, 197–207. Berlin: Mouton de Gruyter.
- TSE, HOLMAN. 2013. Methodological considerations in the study of sociophonetic variation in an underdocumented minority language: Somali Bantu Kizigua as a case study. *Proceedings of the 4th Edition of Journées d'Études Toulousaines (JéTou)*, 129–139. Online: <http://jetou2013.free.fr/documents/JeTou2013-Actes-p129-139-Tse.pdf>
- TUMMERS, JOSE, DIRK SPEELMAN, and DIRK GEERAERTS. 2012. Multiple Correspondence Analysis as heuristic tool to unveil confounding variables in corpus linguistics. In DISTER, ANNE, DOMINIQUE LONGRÉE, and GÉRALD PURNELLE (eds.), *11es Journées internationales d'analyse statistique des données textuelles / Proceedings of the 11th International Conference on the Statistical Analysis of Textual Data*, 923–936. Liège: Université de Liège / Facultés Universitaires Saint-Louis.
- URRY, JAMES. 1991. From speech to literature: Low German and Mennonite identity in two worlds. *History and Anthropology* 5(2). 233–258.
- VAN NESS, SILKE. 1996. Case syncretism in Pennsylvania German: internal or external forces at play? *American Journal of Germanic Linguistics and Literatures* 8(1). 1–17.
- WARKENTIN, KAREN. 2010. 'So ha' wie daut emma jedohne' (*That is how we have always done it*): *the collective memory and cultural identity of the Old Colony Mennonites in Bolivia*. Unpublished MA thesis, University of Manitoba / University of Winnipeg.
- WIEBE, VICTOR. 2013. Gospel Message (radio program). *Global Anabaptist Mennonite Encyclopedia Online*. Online: [http://www.gameo.org/encyclopedia/contents/gospel\\_message\\_radio\\_program](http://www.gameo.org/encyclopedia/contents/gospel_message_radio_program)
- WIELING, MARTIJN, JOHN NERBONNE, and R. HARALD BAAYEN. 2011. Quantitative social dialectology: explaining linguistic variation geographically and socially. *PLoS ONE* 6(9), e23613. doi:10.1371/journal.pone.0023613
- WIENS, CURT. 1916. Niederländischer Einfluß im Wortschatz der Weichselwerder [Netherlandic influence in the lexicon of the Vistula Werders]. *Zeitschrift des Westpreußischen Geschichtsvereins* 56. 139–153.
- WIENS, GERHARD. 1957. Entlehnungen aus dem Russischen im Niederdeutschen der Mennoniten



- in Rußland [Borrowings from Russian in the Low German of the Mennonites in Russia]. *Niederdeutsches Jahrbuch* 80. 93–100.
- WHALEN, DOUGLAS H. 2004. How the study of endangered languages will revolutionize linguistics. In VAN STERKENBURG, PIET (ed.), *Linguistics today: facing a greater challenge*, 321–342. Amsterdam: John Benjamins.
- WOLFRAM, WALT. 1993. Ethical considerations in language awareness programs. *Issues in Applied Linguistics* 4. 225–255.
- WOLFRAM, WALT, JEFFREY REASER, and CHARLOTTE VAUGHN. 2008. Operationalizing linguistic gratuity: from principle to practice. *Language and Linguistics Compass* 2. 1–26.
- WOODBURY, TONY [ANTHONY C.]. 2003. Defining documentary linguistics. In AUSTIN, PETER K. (ed.), *Language Documentation and Description* 1, 35–51. London: Hans Rausing Endangered Languages Project.
- WOODBURY, ANTHONY C. 2011. Language documentation. In AUSTIN, PETER K. and JULIA SALLABANK (eds.), *The Cambridge handbook of endangered languages*, 159–186. Cambridge: Cambridge University Press.
- WURMBRAND, SUSI. 2004. West Germanic verb clusters: the empirical domain. In É. KISS, KATALIN and HENK VAN RIEMSDIJK (eds.), *Verb clusters: a study of Hungarian, German, and Dutch*, 43–85. Amsterdam / Philadelphia: John Benjamins.
- WURMBRAND, SUSI. 2006. Verb clusters, verb raising, and restructuring. In EVERAERT, MARTIN and HENK VAN RIEMSDIJK (eds.), *The Blackwell companion to syntax*, 229–343. Malden, Massachusetts: Blackwell.
- WYNNE, MARTIN (ed.) *Developing linguistic corpora: a guide to good practice*. Oxford: Oxbow Books.
- YAMADA, RACQUEL-MARÍA. 2007. Collaborative linguistic fieldwork: practical application of the empowerment model. *Language Documentation and Conservation* 1(2). 257–282. Online: <http://hdl.handle.net/10125/1717>.
- ZACHARIAS, ED. 2009. *Ons Ieeschtet Wiedabuak* [Our first dictionary]. Altona, Manitoba: Friesens.
- ZIESEMER, WALTHER. 1924. *Die ostpreußischen Mundarten: Proben und Darstellung* [The East Prussian dialects: samples and description]. Breslau: Ferdinand Hirt.

### Appendix A: Mennonite communities in the Saskatchewan Valley

Table 30 and Figure 54 present an overview of early Mennonite settlements in the Saskatchewan Valley established between 1892 and 1908. While many of the Mennonite communities in the Saskatchewan Valley were incorporated as *Strassendorf* villages (a cooperative settlement pattern involving a linear village plan and, often, an open field system of agriculture, developed by Mennonites in northern Poland and Ukraine on the model of earlier northern European village systems; cf. Friesen 1975), others were established instead as so-called ‘four-corner hamlets’ (with families on individual homesteads choosing to build their houses close to one another at the corner of their lands, forming small settlements common across the Canadian prairies; cf. Friesen 1975: 112–113, Guenter et al. 1995) or simply as districts (where settlers’ individual homesteads were concentrated in a particular area).

Several of these communities continue to exist as named entities in the Saskatchewan Valley today: Neuanlage and Neuhorst, for instance, continue to exist as sign-posted hamlets with populations of over one hundred in 2011, while other smaller communities, such as Blumenheim, Blumenthal, and Edenburg, register smaller populations, but continue to receive some official recognition. In general, however, as Guenter et al. (1995) note, many smaller settlements now exist only as historical-geographical reference points or as scattered individual farms, rather than as self-identified communities as such.

<i>Settlement</i>	<i>Type</i>	<i>Est.</i>	<i>Location</i>
Blumenheim	Village	1900	Sec. 31, Tp. 39, R4, W3 (52°23'28.18"N, 106°24'42.78"W)
Blumenhoff	District	1892	Sec. 14 / SW/SE 22, Tp. 42, R3, W3 (52°37'26.56"N, 106°20'23.24"W)
Blumenort	Hamlet / District	1898	Tp. 41, R4, W3 (52°32'36.41"N, 106°33'21.00"W)
Blumenthal	Village	1898	SW 33, Tp. 40, R4, W3 (52°28'55.71"N, 106°22'47.31"W)
Chortitz	Village	1898	SW 5, Tp. 41 / Sec. 32, Tp. 40 (52°29'34.98"N, 106°32'59.02"W)
Clark's Crossing	District	1902	Tp. 37/38, R4/5, W3 (52°16'12.19"N, 106°28'2.96"W)
Edenburg	Village	1902	Sec. 9, 10, 15, Tp. 39, R3, W3 (52°20'50.79"N, 106°21'49.32"W)

<i>Settlement</i>	<i>Type</i>	<i>Est.</i>	<i>Location</i>
Gruenfeld	Village	1899	SW 15, Tp. 40, R5, W3 (52°26'9.30"N, 106°38'33.45"W)
Gruenthal	Village	1898	NE 21, Tp. 40, R4, W3 (52°27'50.21"N, 106°31'11.50"W)
Halbstadt ( <i>Aberdeen</i> )	Hamlet	1901	Sec. 24, Tp. 40, R3, W3 (52°27'24.25"N, 106°18'13.46"W)
Halbstadt ( <i>Hague</i> )	District	1898	Tp. 40/41, R4, W3 (52°29'35.02"N, 106°29'2.00"W)
Hochfeld	Village	1898	NE 15, Tp. 41, R4, W3 (52°31'46.04"N, 106°29'2.07"W)
Hochstadt	Village	1900	NE 26, Tp. 40, R4, W3 (52°28'35.50"N, 106°27'35.56"W)
Hoffnungsort	Hamlet	1899	Tp. 42, R4, W3 (52°36'33.67"N, 106°27'35.58"W)
Krim	District	1908	Tp. 37, R2, W3 (52°10'48.14"N, 106°15'12.66"W)
Kronsthal	Village	1899	Sec. 13/18/19/24, Tp. 39, R3/4, W3 (52°21'43.20"N, 106°26'31.56"W)
Neuanlage	Village	1895	SE 23, Tp. 40, R4, W3 (52°26'57.69"N, 106°27'47.32"W)
Neuhoffnung	Hamlet	1898	Sec. 19, Tp. 41, R3, W3 (52°32'38.33"N, 106°25'25.84"W)
Neuhorst	Village	1898	NW 31, Tp. 39, R4, W3 (52°23'51.02"N, 106°34'47.63"W)
Olgafeld	Hamlet	1902	Sec. 23-26, Tp. 39, R3, W3 (52°22'35.57"N, 106°18'55.45"W)
Osterwick	Village	1899	SW 5 / SE 6, Tp. 39, R4, W3 (52°19'6.49"N, 106°31'54.72"W)
Reinfeld	Village	1897	SW 21, Tp. 41, R3, W3 (52°32'38.38"N, 106°23'16.61"W)
Reinland	Village	1898	NW 23, Tp. 40, R4, W3 (52°25'0.71"N, 106°30'19.80"W)
Rieferthal	District	1905	Tp. 40, R3, W3 (52°24'46.65"N, 106°21'50.04"W)
Rosenbach	Hamlet	1899	Tp. 40, R3, W3 (52°27'23.69"N, 106°21'6.53"W)
Rosenfeld	Village	1902	Sec. 5/8, Tp. 40, R4, W3 (52°25'12.79"N, 106°32'37.38"W)
Rosengart	Village	1899	NE/NW 35, Tp. 41, R4, W3 (52°34'36.07"N, 106°28'19.02"W)
Rosenort	District	1893	Tp. 42, R2, W3 (52°37'52.98"N, 106°11'45.29"W)

<i>Settlement</i>	<i>Type</i>	<i>Est.</i>	<i>Location</i>
Shantzenfeld	Hamlet	1902	Sec. 4, Tp. 39, R2, W3 (52°19'32.14"N, 106°13'54.08"W)
Schlauberg	Hamlet	1908	Sec. 7, Tp. 40, R4/5, W3 (52°25'25.48"N, 106°34'26.30"W)
Schlorrendarp	Hamlet / Village	1905	Sec. 3, Tp. 40, R5, W3 (52°24'46.59"N, 106°38'22.97"W)
Schoenfeld	Hamlet	1902	Sec. 4, Tp. 39, R3, W3 (52°19'32.50"N, 106°22'32.44"W)
Schoenthal	District	1902	Tp. 40, R3/4, W3 (52°25'12.84"N, 106°26'30.36"W)
Schoenwiese	Village	1899	NW 19, Tp. 40, R4, W3 (52°27'49.80"N, 106°34'24.06"W)
Silberfeld	District	1899	Tp. 41, R2/3, W3 (52°30'53.38"N, 106°16'48.13"W)
Steinreich ( <i>Clarkboro</i> )	District	1906	Tp. 38/39, R3, W3 (52°18'13.97"N, 106°24'31.11"W)
Steinreich ( <i>Osler</i> )	District	1906	( <i>n/a</i> )

*Table 30.* Mennonite settlements in the Saskatchewan Valley, 1892–1908 (after Guenter et al. 1995). Includes organized and unorganized villages, four-corner hamlets, and larger districts of Mennonite settlement, with Dominion Land Survey and latitude-longitude coordinates from Guenter et al. (1995) and SaskGrid 2010 GIS datasets.

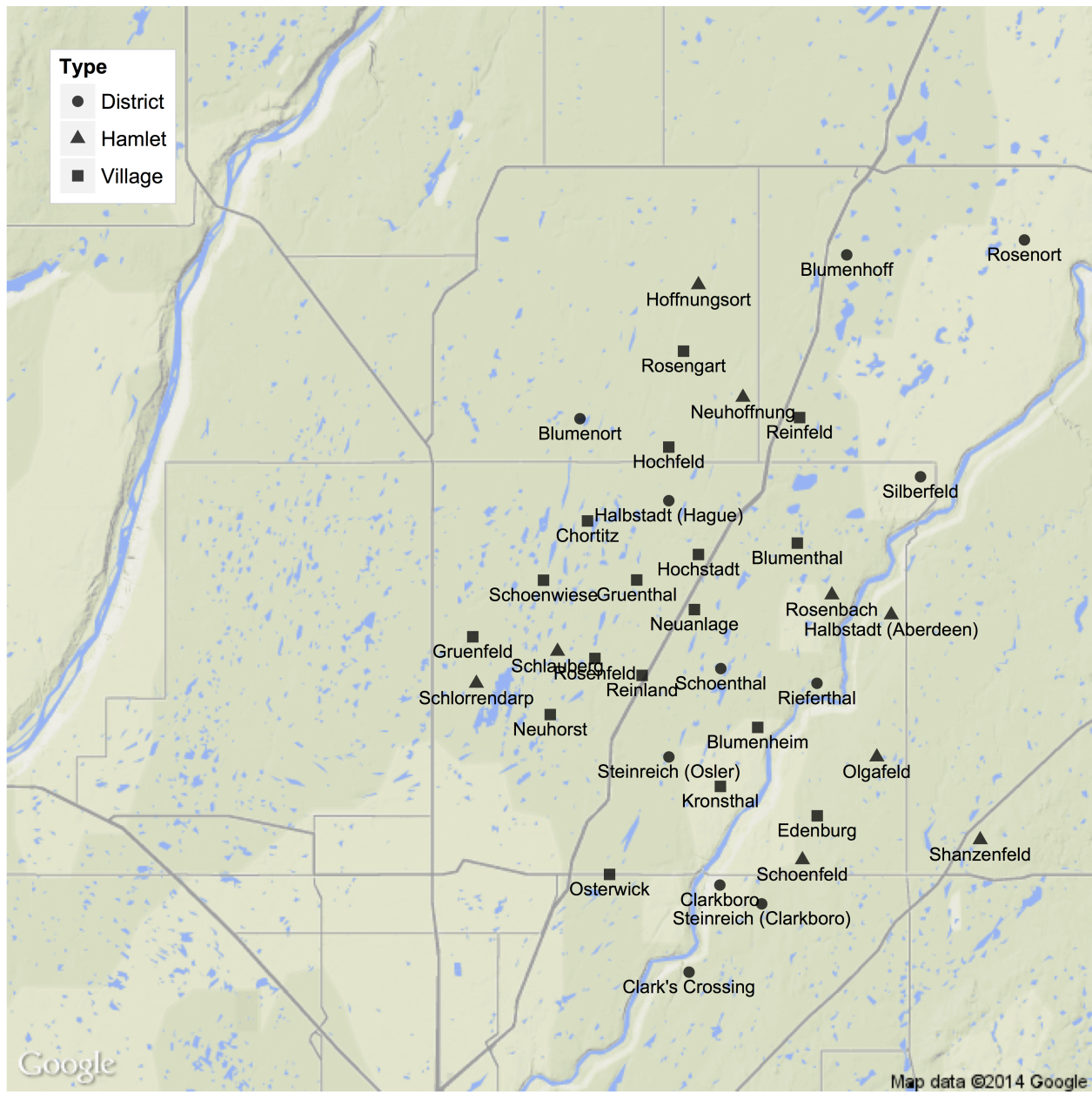


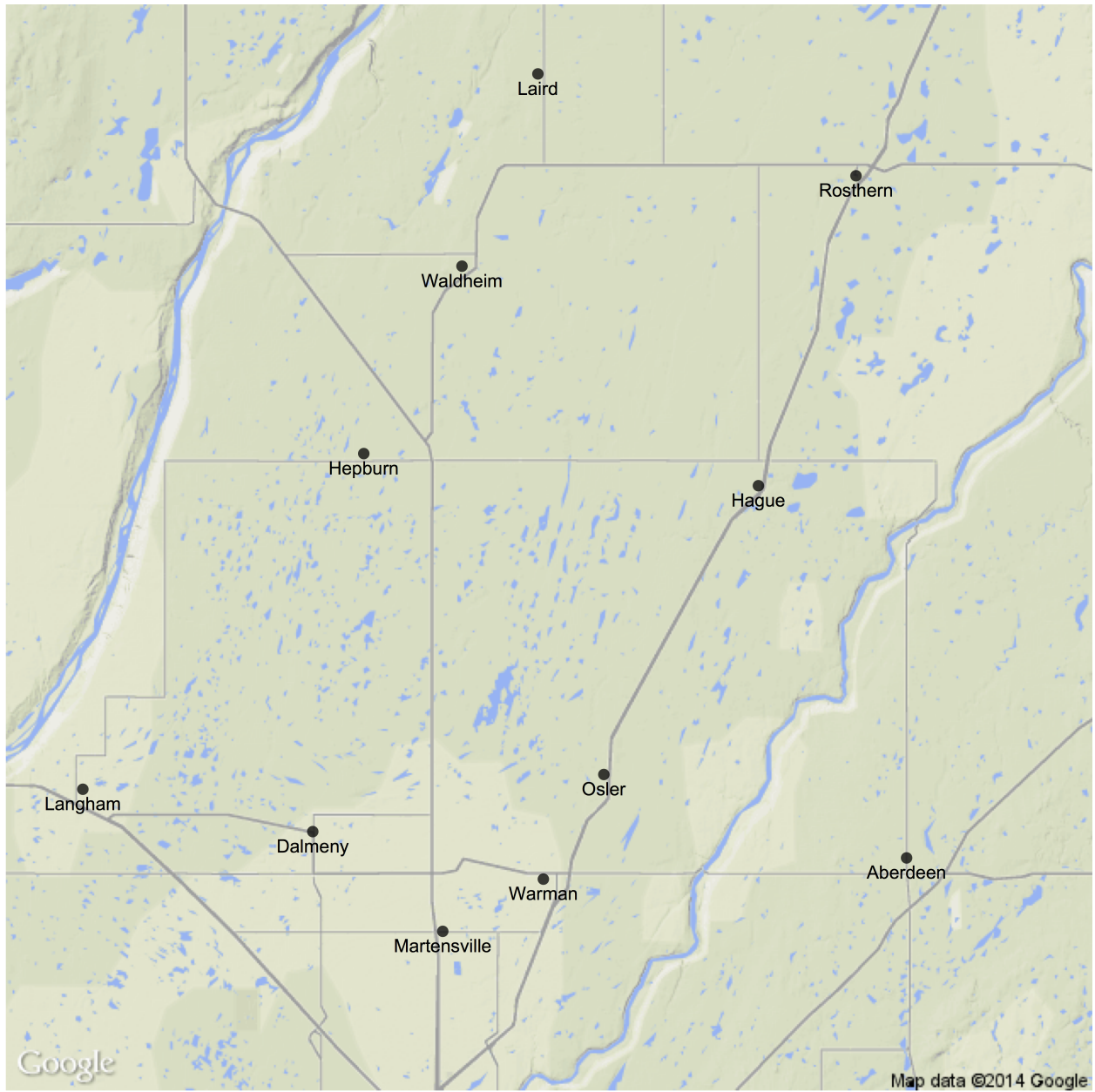
Figure 54. Mennonite settlements in the Saskatchewan Valley, 1892–1908 (after Guenter et al. 1995).

### Appendix B: Non-Mennonite communities in the Saskatchewan Valley

The eleven communities presented in Table 31 and Figure 55 represent important Saskatchewan Valley settlements which were not established primarily by Mennonite immigrants. (Waldheim represents an exception in this respect, with Mennonite settlers in this area of the Saskatchewan Valley first helping to establish the town as a rural centre, and the later arrival of the Canadian North Railway in 1909 contributing substantially to the growth of its non-Mennonite population; see McLennan 2008: 428). Population estimates for each community are drawn from both the 2006 and 2011 Canadian Censuses, demonstrating the recent increase in population in the southern area of the Saskatchewan Valley closest to Saskatoon, which has had a particularly dramatic effect on the towns (now cities) of Martensville and Warman.

<i>Settlement</i>	<i>Est.</i>	<i>Pop. (2006)</i>	<i>Pop. (2011)</i>	<i>Location</i>
Aberdeen	1907	525	599	52°19'33.78"N, 106°17'29.26"W
Dalmeny	1904?	1,515	1,702	52°20'20.60"N, 106°46'18.53"W
Hague	1903	695	878	52°30'35.10"N, 106°24'40.75"W
Hepburn	1919	525	562	52°31'32.00"N, 106°43'50.00"W
Laird	1911	205	287	52°42'43.52"N, 106°35'22.92"W
Langham	1906	1,100	1,290	52°21'35.97"N, 106°57'28.04"W
Martensville	1953	4,965	7,716	52°17'23.00"N, 106°40'0.00"W
Osler	1904	925	1,088	52°22'2.26"N, 106°32'10.87"W
Rosthern	1898	1,355	1,572	52°39'43.61"N, 106°19'56.59"W
Waldheim	1912	800	1,035	52°37'3.84"N, 106°39'3.98"W
Warman	1905	4,730	7,084	52°18'56.00"N, 106°35'7.00"W

*Table 31.* Non-Mennonite-established towns in the Saskatchewan Valley, after Guenter et al. (1995) and McLennan (2008). Population estimates taken from the 2006 and 2011 Canadian Census.



*Figure 55.* Non-Mennonite-established towns in the Saskatchewan Valley, after Guenter et al. (1995) and McLennan (2008).

### Appendix C: Linguistic items

Table 32 provides a summary of variable linguistic items considered in the present study. Such items were identified through review of the relevant literature on dialect variation in Mennonite Plautdietsch speech communities, and through consideration of the responses provided by contributors to the *Fibel* Corpus (cf. §4.2.1).

While the majority of the items presented here were incorporated into the *Fibel* and can be retrieved from responses to the sentences indicated, other items were not explicitly targeted and are thus not consistently available in the corpus. This is the case with several lexical variables, for instance, with items such as ‘nephew’ (LXNEPHEW), ‘rope’ (LXROPE), and ‘either’ (LXEITHER) being wholly unattested, while others such as ‘otherwise’ (LXOTHERWISE) and ‘this (nom. m. sg.)’ (LXTHISNOMMSG) occur only sporadically in contributors’ responses. Several unincorporated items present instances of larger lexical-phonological or morphological phenomena, as with ‘accident’ (LXACCIDENT, targetting the presence or absence of /n/ in the negative prefix *on-* in pre-consonantal environments) and ‘friendly (attrib.)’ (LXFRIENDLY, capturing the inflected forms of *-lich* adjectives as either *-liche*, *-lije*, or *-elje*). In other cases, while a particular item may not be present in the *Fibel*, other items representing the same phenomenon are available: while ‘chew (inf.)’ (LXCHEWINF, with variants *kaue(n)*, *kauwe(n)*, *keiwe(n)*) and ‘sleeve’ (LXSLEEVE, variants *Mau*, *Mauw*, *Meiw*) are not found in the *Fibel* Corpus, items such as ‘blue’ (LXBLUE, variants *blau*, *blauw*, *bleiw*) and ‘grey’ (LXGREY, variants *grau*, *grauw*, *greiw*) are robustly attested, and have widely been claimed to reflect the same pattern.<sup>58</sup> While unattested items are in the minority, their absence from the corpus nevertheless leaves unfortunate gaps in the description of variation in Mennonite Plautdietsch. These items have been documented in this appendix alongside other, more consistently represented forms, in the hope of facilitating further investigation into this additional variation in the future.

---

<sup>58</sup> While the presence of such ‘surrogate’ items is no doubt useful, it cannot be assumed that all items associated with a particular lexical-phonological or morphological phenomenon necessarily pattern in exactly the same way, and that other, undocumented items are therefore redundant. As lexically exhaustive coverage of a particular phenomenon is typically not feasible, however, the inclusion of multiple items hypothesized to instantiate a particular pattern would appear to present a reasonable compromise, given the aims of this study.



#	Category	Item	Variants	Identifier	Sources
1	LEX	‘a (nom. m./n.)’	<i>e’, een, ‘en</i>	LXMASCNEUTA	S14, 21, 32
2	LEX	‘am’	<i>se’, senn, si</i>	LXAM	S18
3	LEX	‘any’	<i>irjend, injend</i>	LXANY	( <i>n/a</i> )
4	LEX	‘are’	<i>sen’, send, senne</i>	LXARE	S04, 23, 46
5	LEX	‘as’	<i>aus, auls</i>	LXAS	( <i>n/a</i> )
6	LEX	‘aunts’	<i>Tauntes, Mumms</i>	LXAUNTS	S23
7	LEX	‘because’	<i>wääjens, wiel(s), wielt, ...</i>	LXBECAUSE	S08,16,24,32,37,40,48
8	LEX [PHON,MORPH]	‘been’	<i>jewas(t), jewäse(n)</i>	LXBEEN	S11, 14
9	LEX	‘between’	<i>teschen, tweschen</i>	LXBETWEEN	S15
10	LEX	‘bird (gender)’	MASC., NEUT.	LXBIRDGENDER	S19
11	LEX	‘down, off of’	<i>(e)rauf, (e)raufa</i>	LXDOWN	S39
12	LEX	‘early’	<i>tiedig, fräh</i>	LXEARLY	S31
13	LEX [MORPH]	‘eaten’	<i>jeäte(n), jejäte(n)</i>	LXEATEN	S28
14	LEX	‘either’	<i>entwäda, entswäda</i>	LXEITHER	( <i>n/a</i> )
15	LEX	‘English’	<i>Engelsch, Englisch</i>	LXENGLISH	S19
16	LEX [PHON]	‘farmer’	[fɔrma], [bua], ...	LXFARMER	S16, 36
17	LEX	‘George’	<i>Jeat, Jorg</i>	LXGEORGE	S40
18	LEX [MORPH]	‘girls’	<i>Me(r)jalle(n), Me(r)jalles, Määkjjes</i>	LXGIRLS	S01, 24
19	LEX	‘grandmother’	<i>Groosma(u), Grootmutta, ...</i>	LXGRANDMOTHER	S55
20	LEX [MORPH]	‘her (dat.)’	<i>äh, äaht, ahr</i>	LXDATHER	S03, S52
21	LEX	‘immediately’	<i>fuat, fuats, soofuat</i>	LXIMMEDIATELY	( <i>n/a</i> )
22	LEX	‘into’	<i>(e)nenn, (e)nenna</i>	LXINTO	S39
23	LEX	‘knew’	<i>wisst, wusst</i>	LXKNEW	S06, 20
24	LEX	‘lap (gender)’	FEM., MASC.	LXLAPGENDER	( <i>n/a</i> )
25	LEX	‘little’	<i>kjlien, kjleen</i>	LXLITTLE	S01
26	LEX	‘nephew’	<i>Neffe, Plemmenikj, Sobrino, Vada</i>	LXNEPHEW	( <i>n/a</i> )
27	LEX [PHON,MORPH]	‘often’	[foəkə], [føøkən], [ɔft], ...	LXOFTEN	S29
28	LEX	‘otherwise’	<i>sessst, sonst</i>	LXOTHERWISE	( <i>n/a</i> )
29	LEX [PHON]	‘out’	[ərut], [əruta], [ryt], [ryta], ...	LXOUT	S08
30	LEX	‘rope’	<i>Knaut, Strang</i>	LXROPE	( <i>n/a</i> )
31	LEX	‘rub’	<i>rubble(n), schobbe(n), ...</i>	LXRUB	S49
32	LEX [MORPH]	‘say (inf.)’	<i>saije(n), saje(n)</i>	LXSAY	S17, 51
33	LEX [PHON]	‘seventy-four’	<i>vea(r)u(n)säwentig, -zäwentig</i>	LXSEVENTYFOUR	( <i>n/a</i> )
34	LEX	‘some’	<i>atlije, some, walkje, waut, ...</i>	LXSOME	S03
35	LEX	‘store’	<i>Laufkje, Lode(n), Stua</i>	LXSTORE	S54

#	Category	Item	Variants	Identifier	Sources
36	LEX [MORPH]	‘them (dat.)’	<i>ähnt, ahn</i>	LXDATTHEM	S28
37	LEX	‘this (nom. m. sg.)’	<i>dis’, disa, diss’, dissa</i>	LXTHISNOMMSG	( <i>n/a</i> )
38	LEX	‘uncles’	<i>Onkels, Oohms</i>	LXUNCLES	S23
39	LEX	‘under’	<i>inja, unja</i>	LXUNDER	( <i>n/a</i> )
40	LEX	‘until’	<i>bat, bott</i>	LXUNTIL	S29
41	LEX	‘watermelon’	<i>Arbus, Rebus, Wotameloon, ...</i>	LXWATERMELON	S53
42	LEX	‘whether’	<i>auf, aus, es</i>	LXWHETHER	S19, 36, 46
43	LEX [PHON]	‘would (2s.)’	<i>wuddst, wu(r)scht</i>	LXWOULD2S	S47
44	LEX [MORPH]	‘you (acc. pl.)’	<i>ju, junt</i>	LXYOUACCP	( <i>n/a</i> )
45	LEX-PHR	‘at the beginning’	<i>aum/em Au(n)fang, ...</i>	CXATTHEBEGINNING	S51
46	LEX-PHR [MORPH]	‘every year’	<i>aula Joah, jieda Joah, ...</i>	CXEVERYYEAR	S25
47	LEX-PHR	‘in the evening’	<i>em/opp’en Owend, zeowenst, ...</i>	CXINTHEEVENING	S31
48	LEX-PHR [MORPH]	‘into the house’	<i>em/en daut Huus (‘enenn), ...</i>	CXINTOTHEHOUSE	S39
49	LEX-PHR [MORPH]	‘off of the wagon’	<i>vom/von dän W. (‘erauf), ...</i>	CXOFFOFTHEWAGON	S39
50	LEX-PHR	‘that (indef. rel. clause)’	<i>daut, waut</i>	CXINDEFRELCLAUSE	S09
51	LEX-PHR	‘that (n. rel. clause)’	<i>daut, waut, woont</i>	CXNEUTERRELCLAUSE	S52
52	LEX-PHR	‘that (pl. rel. clause)’	<i>daut (doa), dee, waut (doa), ...</i>	CXPLURALRELCLAUSE	S34
53	LEX-PHR [MORPH]	‘that day’	<i>dee/däm/dän Dag</i>	CXTHATDAY	S08
54	LEX-PHR [MORPH]	‘the one (focus)’	<i>dee/däm/dän (eenzja, ...), ...</i>	CXTHEONE	S18
55	LEX-PHR [MORPH]	‘without’	<i>ohne(n), met ohne(n)</i>	LXWITHOUT	S35
56	LEX-PHON	‘accident’	<i>Onjlekj, O’jlekj</i>	LXACCIDENT	( <i>n/a</i> )
57	LEX-PHON	‘and’	[æ̃n], [ɛ̃n], [ə̃n], [ʌ̃n], [ʊ̃n]	LXAND	S23, 39, 41
58	LEX-PHON [MORPH]	‘ate (pl.)’	<i>aute(n), eete(n)</i>	LXATE	S35
59	LEX-PHON	‘blue’	<i>blau, blauw, bleiw</i>	LXBLUE	S03, 46
60	LEX-PHON	‘came (sg.)’	<i>kaum, kjeem</i>	LXCAMESG	( <i>n/a</i> )
61	LEX-PHON	‘can (2s.)’	<i>kaunst, kau’st</i>	LXCAN2S	S31
62	LEX-PHON [MORPH]	‘can (pl.)’	<i>kjänne(n), kjenne(n)</i>	LXCANPL	S03, 05
63	LEX-PHON	‘can you (pl.)’	<i>kjä(nn)’ (ji), kje(nn)’ (ji), ...</i>	LXCANYOUP	S03, 31
64	LEX-PHON [MORPH]	‘chew (inf.)’	<i>kaue(n), keiwe(n)</i>	LXCHEWINF	( <i>n/a</i> )
65	LEX-PHON	‘could (2s.)’	<i>kunnst, ku’st</i>	LXCOULD2S	S17
66	LEX-PHON [MORPH]	‘gave (pl.)’	<i>gauwe(n), jeewe(n)</i>	LXGAVE	S43
67	LEX-PHON	‘grey’	<i>grau, grauw, greiw, jreiw</i>	LXGREY	S19, 37, 41
68	LEX-PHON [MORPH]	‘have (aux. inf.)’	<i>habe(n), ha’(n), hawe(n), ...</i>	LXHAVEINF	S47
69	LEX-PHON [MORPH]	‘have (aux. pl.)’	<i>habe(n), ha’(n), hawe(n), ...</i>	LXHAVEAUXPL	S29
70	LEX-PHON [MORPH]	‘have (lex. pl.)’	<i>habe(n), ha’(n), hawe(n), ...</i>	LXHAVELEXPL	S09, 30, 41

#	Category	Item	Variants	Identifier	Sources
71	LEX-PHON	‘horses’	[p <sup>h</sup> eəd], [p <sup>h</sup> iəd]	LXHORSES	S41
72	LEX-PHON	‘sat (sg.)’	<i>saut, seet</i>	LXSAT	S10
73	LEX-PHON [MORPH]	‘shall (pl.)’	<i>sälle(n), selle(n)</i>	LXSHALLPL	(n/a)
74	LEX-PHON	‘should (2s.)’	<i>sullst, su’st</i>	LXSHOULD2S	S07
75	LEX-PHON	‘sleeve’	<i>Mau, Mauw, Meiw</i>	LXSLEEVE	(n/a)
76	LEX-PHON	‘supper’	[ovənkəs(t)], [ʊŋkəs], ...	LXSUPPER	S13
77	LEX-PHON [MORPH]	‘took (pl.)’	<i>nauhme(n), neehme(n), noohme(n)</i>	LXTOOK	S43
78	LEX-PHON	‘was’	[veə], [via]	LXWAS	S16, 42, 52
79	PHON [MORPH]	‘become (part.)’	[jəvɔɪdə(n)], [jəvɔrdə(n)]	LXBECAME	S36, 37, 45, 50
80	PHON [LEX]	‘became (sg.)’	[vɔɪd], [vɔɪt <sup>h</sup> ], [vɔrt <sup>h</sup> ]	LXBECAMESG	S36, 37, 45, 50
81	PHON [MORPH]	‘berries’	[beəɪn], [beərə(n)], ...	LXBERRIES	S03
82	PHON [MORPH]	‘cook (inf.)’	[keʊkə(n)], [koəkə(n)], ...	LXCOOK	S13, 36
83	PHON	‘cooked (part.)’	[jəkoəkt], [jəkøʊkt], ...	LXCOOKED	S36
84	PHON [LEX]	‘corn, maize’	[k <sup>h</sup> ɔɪn], [k <sup>h</sup> ɔɪn], [k <sup>h</sup> ukəʊz], ...	LXCORN	S15
85	PHON [LEX]	‘days’	[doːʊə], [doəg], [deʊy], ...	LXDAYS	S25, 08
86	PHON [MORPH]	‘make (inf.)’	[meʊkə(n)], [moəkə(n)], ...	LXMAKE	S45, 13
87	PHON [MORPH]	‘parents’	[elərə(n)], [el(d)əɪn]	LXPARENTS	S15, 43
88	PHON	‘sixty-one’	[eːnənsastɪç], [eːnənʃastɪç], ...	LXSIXTYONE	S09
89	PHON [MORPH]	‘sugar cookies’	[ʃʊkakuəkə], [sɔkakyəkən], ...	LXSUGARCOOKIES	S47
90	PHON [LEX]	‘today’	[fəndəʊg], [fəneʊən], ...	LXTODAY	S05, 13
91	MORPH	‘big (acc. m. def.)’	<i>groota, grooten</i>	LXMASCACCDEFBIG	S07
92	MORPH	‘brown (acc. m. def.)’	<i>bruuna, bruunen</i>	LXMASCACCDEFBROWN	S07
93	MORPH	‘drive (inf.)’	<i>foahre, foahren</i>	LXDRIVEINF	S17
94	MORPH	‘friendly (attrib.)’	<i>frindliche, frintlīje, frintelje</i>	LXFRIENDLY	(n/a)
95	MORPH	‘given’	<i>jejäwt, jejäwe(n)</i>	LXGIVEN	S42
96	MORPH	‘ham’	<i>Schinkjefleesch, Schinkjenfleesch</i>	LXHAM	(n/a)
97	MORPH	‘hams’	<i>Schinkjens, Schinkjes</i>	LXHAMS	S13
98	MORPH	‘helped (past pl.)’	<i>halpde(n), holpe(n)</i>	LXHELPED	S55
99	MORPH	‘our (dat. m. sg.)’	<i>ons, onsem, onsen, ...</i>	LXMASCDATOUR	S11
100	MORPH [PHON]	‘painted (part.)’	<i>jefoawt, jeforwe(n), ...</i>	LXPAINTED	S10
101	MORPH	‘Peter (acc.)’	<i>Peetren, Peeta</i>	LXPETERACC	(n/a)
102	MORPH	‘ <i>Plume(n)moos</i> (soup)’	<i>Plumemoos, Plumenmoos</i>	LXPLUMMOOS	(n/a)
103	MORPH	‘rabbit (acc.)’	<i>Hosen, Hos</i>	LXRABBIT	(n/a)
104	MORPH	‘read (part.)’	<i>jeläst, jeläse(n)</i>	LXREADPART	(n/a)
105	MORPH	‘sing (inf.)’	<i>sinje, sinjen</i>	LXSINGINF	S06

#	Category	Item	Variants	Identifier	Sources
106	MORPH [PHON]	‘soft (n. indef.)’	<i>wäkjet, weakja, weakjet</i>	LXNEUTINDEF_SOFT	S21
107	MORPH	‘the (acc. m. sg.)’	<i>däm, dän, de</i>	LXMASCACCTHE	S07, 18
108	MORPH	‘the (dat. m. sg.)’	<i>däm, dän</i>	LXMASCDATTHE	S17
109	MORPH	‘the (m. sg. possessor)’	<i>dee, däm, dän</i>	CXMASCPOSSTHE	S36
110	MORPH	‘them (dat. def.)’	<i>dän, dee</i>	LXDATTHEMDEF	S28
111	MORPH	‘they (enclitic) had’	<i>haude se, hauden se</i>	LXTHEYENCLITICHAD	(n/a)
112	MORPH	‘they (enclitic) said’	<i>säde se, säden se</i>	LXTHEYENCLITICSAID	(n/a)
113	MORPH	‘visited (part.)’	<i>spazeat, jespazeat</i>	LXVISITED	S29
114	MORPH	‘wagons’	<i>Woagens, Woages</i>	LXWAGONS	S41

*Table 32.* Variable linguistic items by category (LEX: lexical selection involving single lexemes, LEX-PHON: lexically restricted phonological patterns, LEX-PHR: lexical selection involving phrases and multi-word expressions, PHON: general phonological patterns, MORPH: morphological patterns (case inflection, paradigm selection)). Items are given with attested variants (in the Sass orthography or in IPA, according to the corpus coding conventions for each item), as well as identifiers and corresponding sentences in the Fibel Corpus on which items are coded.

### **Appendix D: Contributor information forms**

The following three-page form was provided to contributors to the *Fibel* Corpus to gather information about their linguistic, personal, and family background. The personal and family information sections reproduced in Figure 56 concentrate on basic sociodemographic details (e.g., the age, gender, occupational and educational background, denominational affiliation, and residential history of contributors and their parents), while the sections on linguistic background in Figure 57 and Figure 58 explore in greater detail patterns of language use (e.g., typical interlocutors and contexts of use for each language, general frequency of language use) and self-assessed proficiency for each of the languages with which the contributor has some familiarity.



# Onse Sproak • “Our Language”

Community-partnered documentation of Plautdietsch

This short form is meant to include information to help (a) learn more about how people of different backgrounds speak Plautdietsch, and (b) make sure that all contributions are correctly acknowledged. This information will only ever be shared generally in an anonymized form.

## 1. Personal information

Full name: \_\_\_\_\_ Gender: M / F

Place of birth: \_\_\_\_\_ Date of birth: \_\_\_\_ / \_\_\_\_ / \_\_\_\_

Occupation (if retired, please note earlier occupation(s)): \_\_\_\_\_

Church congregation affiliation (if any): \_\_\_\_\_

Where have you lived before? (for 3+ years): \_\_\_\_\_

\_\_\_\_\_

Education (please circle): [ Before grade eight – After grade eight – High school – Professional training – University ]

## 2. Family information

Father's full name: \_\_\_\_\_

Place of birth: \_\_\_\_\_ Date of birth: \_\_\_\_ / \_\_\_\_ / \_\_\_\_

Occupation: \_\_\_\_\_

Church congregation affiliation (if any): \_\_\_\_\_

Mother's full name (with maiden name): \_\_\_\_\_

Place of birth: \_\_\_\_\_ Date of birth: \_\_\_\_ / \_\_\_\_ / \_\_\_\_

Occupation: \_\_\_\_\_

Church congregation affiliation (if any): \_\_\_\_\_

Figure 56. Contributor information form, personal and family information.

### 3. Language information

Please list all of the **languages you know** in the order you learned them (your native language first):  
 (if you speak more than one variety or dialect, please list each one separately – e.g. “Low German” and “High German”, not just “German”)

List languages here →	A.	B.	C.	D.
-----------------------	----	----	----	----

**How often** do you use each of these languages?  
 (for each language, please circle on a scale of zero to three, where 0 = *never*, 1 = *less than once a week*, 2 = *more than once a week*, 3 = *every day*)

List languages here →	A.	B.	C.	D.
How often →	0   1   2   3	0   1   2   3	0   1   2   3	0   1   2   3

**How well** would you say you can **understand**, **speak**, and **read** in each of these languages?  
 (for each language, please circle on a scale of zero to five, where 0 = *not at all*, 1 = *a little bit*, 2 = *passably*, 3 = *well*, 4 = *very well*, 5 = *fluently*)

List languages here →	A.	B.	C.	D.
<b>Understanding:</b>	0   1   2   3   4   5	0   1   2   3   4   5	0   1   2   3   4   5	0   1   2   3   4   5
<b>Speaking:</b>	0   1   2   3   4   5	0   1   2   3   4   5	0   1   2   3   4   5	0   1   2   3   4   5
<b>Reading:</b>	0   1   2   3   4   5	0   1   2   3   4   5	0   1   2   3   4   5	0   1   2   3   4   5

Figure 57. Contributor information form, language proficiency information.

Who do you speak each of these languages with?

(for each language, please circle from zero to five, where 0 = never, 1 = very rarely, 2 = rarely, 3 = occasionally, 4 = very frequently, 5 = always, or n/a = not applicable)

List languages here →	A.	B.	C.	D.
With my <b>grandparents</b> (when they were alive)	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
With my <b>parents</b> (when they were alive)	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
With my <b>spouse</b> (when s/he was alive)	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
With my <b>siblings</b> (when they were alive)	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
With my <b>children</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
With my <b>extended family</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
With my <b>friends</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
With my <b>neighbours</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
With my <b>co-workers</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
With <b>strangers</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a

Where do you use each of these languages?

(for each language, please circle from zero to five, where 0 = never, 1 = very rarely, 2 = rarely, 3 = occasionally, 4 = very frequently, 5 = always, or n/a = not applicable)

List languages here →	A.	B.	C.	D.
At <b>home</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
At <b>friends' houses</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
At the <b>coffee shop</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
At <b>church</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
At <b>work</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
At <b>school</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
At <b>public events</b> (auctions, weddings, etc.)	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
When <b>writing</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a
When <b>praying</b>	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a	0   1   2   3   4   5   n/a

Figure 58. Contributor information form, language use information.



## Appendix E: Transcription conventions

The transcription conventions described in this appendix were largely adapted from other, published transcription systems in corpus linguistics and sociolinguistics. In particular, these conventions draw heavily on the recommendations of the International Corpus of English (ICE; Nelson 2002) and Santa Barbara Corpus of Spoken American English (Du Bois et al. 2000–2005) for the representation of spoken texts. For example, the short and long pause notations [.] and [..] used here are adapted from the equivalent sequences <,> and <,,> in the ICE conventions. The use of [anon descr="..."] ... [/anon] markers to indicate sections of speech requiring anonymization is similar to the ICE conventions <@> ... </@> for changed names and words, but preserve the original text of the elements to be anonymized. Other conventions are more typical across transcription systems: the representation of truncated words in the *Fibel* Corpus with a trailing hyphen follows both the practices of the Santa Barbara Corpus of Spoken American English and several systems described elsewhere (e.g., Tagliamonte 2006).

These existing transcription conventions were adopted selectively, rather than wholesale, for several reasons. As both the ICE and Santa Barbara systems define extensive, detailed sets of transcription symbols for features of spoken language (e.g., counting the number of pulses of laughter, or indicating ‘smiled’ speech), selecting a smaller number of these conventions allowed for transcription to proceed more rapidly without abandoning all features of interest. Moreover, adapting these systems provided an opportunity to ensure that the character sequences used to encode their conventions did not conflict with the other forms of representation for digital text used in this project. While many of the original ICE transcription conventions were represented with SGML tags (such as <,> and <,,> noted above), several of these symbols do not conform with the XML standard used elsewhere for texts in this corpus, and could potentially raise difficulties for later computational processing of transcribed documents. Similarly, adaptations sometimes allowed for more precise descriptions of particular spoken language phenomena. In contrast with the system advanced by Tagliamonte (2006), where false starts are indicated only on the final word of a longer sequence (and thus require human readers to interpret where the fragment begins), these conventions enclose the entire sequence of words associated with a false start in [fs] ... [/fs] tags, making them accessible to automatic processing later on. Adapting these published conventions thus attempted to strike a balance between common practices in

linguistic transcription and the need to ensure that the final conventions were circumscribed, consistent, and amenable to computational processing. The final set of conventions is summarized in Table 33, with their use illustrated with examples from the *Fibel* Corpus in (23a–n) below.<sup>59</sup>

<i>Convention</i>	<i>Interpretation</i>
xyz-	Hesitation, truncated word
abc-- // --def	Word continuation across pause boundary
[.]	Short pause (≤ one syllable in duration)
[..]	Longer pause (> one syllable in duration)
[anon descr=“...”] ... [/anon]	Sections to be omitted when anonymized
[e] ... [/e]	Speech in English
[emph] ... [/emph]	Stressed, emphatic speech ( <i>optional</i> )
[event descr=“...” /]	Non-speech event (e.g., cough, phone ringing)
[ext] ... [/ext]	Perseveration, extended speech duration ( <i>optional</i> )
[fs] ... [/fs]	False start within a larger utterance
[lang id=“...”] ... [/lang]	Speech in a language other than Plautdietsch
[laugh /]	Laughter
[u] ... [/u]	Unclear speech, uncertain transcription
[u /]	Unclear speech, no transcription attempted
[voice quality=“...”] ... [/voice]	Voice quality (e.g., whisper, laughing)

Table 33. Transcription conventions in the *Fibel* Corpus.

- Hesitation or truncated word*
- (23) a. Oh, ji sullen dän **bru-**, [.] grooten, bruunen Boa auleen loten.  
 oh you should the HES big brown bear alone let. INF  
 ‘Oh, you should leave the br-, big brown bear alone.’  
 (F18, 2011-08-04, 04m28s630–04m32s630)
- Word continuation across pause boundary*
- b. Dis' [.] oola Maun bädt, daut siene **Groot-- // --kjinja //** hawe goot jeschlope.  
 this old man prays that his grand-- --children have well slept  
 ‘This old man is praying that his grandchildren have slept well.’  
 (M05, 2011-08-04, 00m31s190–00m36s620)

<sup>59</sup> Several of these conventions serve as shorthand representations of situations or events that could also be represented in other ways under this system. Thus, marking off sections of speech that appear in English with [e] ... [/e] could also be achieved equivalently with [lang id=“eng”] ... [/lang], using the standard ISO 639-3 code for English (eng) in the [lang] ... [/lang] tag. Likewise, [laugh /] offers a shortened form of [event descr=“laugh” /], which serves the same purpose. While not strictly necessary, conventions such as these allow for more parsimonious representations of speech events that occur frequently in the corpus.

*Short pause*

- c. Dee haft 'en weakjet [.] Hoat.  
he has a soft heart  
'He has a soft heart.'  
(F03, 2011-09-13 (01), 14m38s170–14m40s390)

*Long pause*

- d. Um, he es // weakj[.]hoatig.  
um he is soft-hearted  
'Um, he is soft-hearted.'  
(M04, 2011-08-02, 08m31s600–08m35s866)

*Anonymized speech*

- e. [e]See,[/e] Taunte [anon descr="name"]XXX[/anon] wudd saije: "Goaden."  
see aunt XXX would say garden  
'See, Aunt XXX would say Goaden.' (XXX = name)  
(F06, 2012-10-18, 24m01s985–24m03s705)

*Speech in English*

- f. [e]"Rub" is[/e] "schobbe."  
rub is rub.INF  
""Rub" is schobbe."  
(F06, 2012-10-18, 53m03s200–53m04s350)

*Stressed, emphatic speech*

- g. ...wiels hia noch emma Dietsch jerädt [emph]woat[/emph].  
because here still always German spoken becomes  
'...because German is still spoken here.' (contrasting with preceding was)  
(M16, 2011-07-23, 00m33s390–00m36s430)

*Non-speech event*

- h. [event descr="clear throat" /] Wöagen.  
'(clears throat) Wagon.'  
(F15, 2011-10-26, 30m18s461–30m19s160)

*False start*

- i. Daut kaun eena goanich [fs]opp enjle-,[/fs] opp Dietsch saijen,  
that can one not.at.all in Engli- in German say. INF  
'You can't say that at all in Engli-, in German.'  
(M04, 2011-08-02, 08m26s776–08m29s206)

*Non-Plautdietsch speech*

- j. [lang id="deu"]Laufen[/lang], rannen, nich?  
Laufen run.INF not  
'Laufen, to run, right?' (laufen 'to run', Standard German (ISO 639-3: deu))  
(M18, 2011-08-02, 50m19s885–50m21s175)

*Laughter*

- k. Eena haud doa uk kunnt saijen, "Jeat sien Knoss wea soo..." [laugh /]  
one had there also could say.INF George his animal.foot was so  
'You could've also said, "George's foot (vulg.) was so..." (laughs)'  
(M17, 2011-10-29, 40m32s840–40m35s160)

*Unclear speech, uncertain transcription*

- l. ...daut es äwajens nich soo's 'et hia [u]han[/u] jeschräwen es.  
that is besides not so-as it here thither written is  
'...anyway, that's not how it's written (down?) here.'  
(F02, 2011-08-03 (02), 08m42s170-08m43s830)

*Unclear speech, no transcription attempted*

- m. ...daut [u /] eascht jeköakt word.  
that first boiled became  
'...that (?) got boiled first.'  
(M14, 2011-08-03 (02), 12m16s365-12m17s685)

*Voice quality*

- n. [voice quality="whisper"]Wi hawe foaken met onse Nobasch spazeat[/voice]  
we have often with our neighbours visited  
'(whispering) We have often visited with our neighbours'  
(F03, 2011-09-13 (01), 21m40s440-21m43s830)